History-matching of petroleum reservoir models by the ensemble Kalman filter and parameterization methods

Leila Heidari

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

Ecole doctorale n° 432 : Sciences des Métiers de l' Ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

" Spécialité "Géostatistiques "

présentée et soutenue publiquement par

Leila HEIDARI

21 janvier 2011

History-Matching of Petroleum Reservoir Models by the Ensemble Kalman Filter and Parameterization Methods

Calage d'Historiques de Réservoirs Pétroliers par le Filtre de Kalman d'Ensemble et des Méthodes de Paramétrisation

Directeur de thèse : **Hans WACKERNAGEL**

Co-encadrement de la thèse : **Véronique GERVAIS-COUPLET**

Jury

M. Geir Evensen, Statoil et Université de Bergen	Rapporteur
M. Dominique Gibert, Institut de Physique du Globe (IPG) de Paris	Rapporteur
Mme. Véronique Gervais-Couplet, IFP Energies nouvelles, Maître de Thèse	Examinatrice
Mme. Mickaele Le Ravalec-Dupin, IFP Energies nouvelles	Examinatrice
M. Richard Rivenq, Total E&P, Aberdeen	Examinateur
M. Hans Wackernagel, MINES ParisTech, Directeur de Thèse	Examinateur

T H E S E

MINES ParisTech

Centre de Géostatistique

35 Rue St Honoré 77300 Fontainebleau

A

Mon pays, l’Iran

Mes parents, Ghasem et Parvin

Ma soeur, Parisa

Acknowledgments

This research project would have not been realized without the people contributed with their expertise and precious help.

First and foremost, I have to express my sincere gratitude to my thesis supervisor Dr. Hans Wackernagel for helping me to develop an accurate understanding of a research project and providing me with the means to pursue such an act.

I am absolutely thankful to my thesis advisor Dr. Véronique Gervais-Couplet, for her patience, continuous support and insightful comments during the course of this research. Her availability for discussions and her encouragements for trying new solutions when the project got blocked was one of the main sources of motivation in these three years.

Equally, I would like to thank my thesis co-advisor Dr. Mickaele Le Ravalec-Dupin for providing me the opportunity to perform a thesis at IFP Energies nouvelles. I appreciate her presence for solving the encountered problems, despite her busy time schedule. Her passionate and mindful way of attacking new ideas made me familiar with the basics of a research work.

I would like to thank the jury members, Dr. Geir Evensen, Dr. Richard Rivenq and Dr. Dominique Gibert for their time and consideration for this thesis.

The assistance received from the “Fluid Flow Department” of IFP Energies nouvelles and the “Geostatistics Group” of MINES-ParisTech is of invaluable importance. I offer my regards to Dr. Frederic Roggero, Dr. Patrick Lemonnier and Dr. Jean-Paul Chilès. Special thanks to Benoit Noetinger, Lin. Y. Hu, Guillaume Enchery, Sébastien Da Vega, Elodie Tillier, Sylvie Hoguet, Nathalie Dietrich and Dominique Vasiliadis.

I am indebted to my many student colleagues for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Setayesh Zandi, Samir Touzani, Jeremy Rosak, Ekaterina Sergienko, Zyed Bouzarkouna and Marius Verscheure. Meanwhile, special thanks to Mandana, Imen, Shora and Azin, who lived at the same time the same experience as mine, for their kind and friendly support.

Last but not least, I thank my parents, Ghasem and Parvin, and my sister, Parisa, for their guidance, support and love all throughout my life without which my life was meaningless.

Résume

Le calage historique est une étape essentielle dans toute étude de réservoir. Il permet l'intégration de données acquises après la production dans la construction de modèles de réservoir. Il existe plusieurs méthodes pour effectuer un tel acte. Le filtre de Kalman d'Ensemble (EnKF) est une méthode d'assimilation (ou calage historique) séquentielle capable d'intégrer les données mesurées dès qu'ils sont obtenus. Plusieurs applications réussies de la méthode sont connues dans les applications océanographiques, météorologiques et du pétrole. Toutefois, la performance de la méthode peut se détériorer lorsque la taille de l'ensemble est petite ou lors du traitement des modèles non Gaussiennes ou fortement non linéaire.

Ce travail est basé sur l'application de l' EnKF pour le calage historique et est divisé en deux sections principales.

La première section traite l'application de la EnKF à plusieurs cas d'études afin de mieux comprendre les avantages et les inconvénients de la méthode. Ces cas d'étude incluent deux cas d'étude synthétiques (un simple et un plutôt complexe), un modèle de faciès et un modèle de réservoir réel. Dans la plupart des cas, la méthode a réussi à reproduire les données mesurées. Les problèmes rencontrés, tels que la surestimation des paramètres pour le cas assez complexe, sont expliqués et des solutions possibles sont proposées.

La seconde partie traite deux nouveaux algorithmes proposé en combinant l'EnKF avec deux méthodes de paramétrisation: méthode des points pilotes et méthode de déformation graduelle. Les deux méthodes sont capables de préserver les propriétés statistiques de l'ordre de deux (moyenne et covariance). Leur intégration à l'EnKF assure la conservation des propriétés a priori de modèle au cours du processus d'assimilation. Les deux algorithmes développés sont appliqués au cas d'étude synthétique simple. L'application a été réussit pour la méthode des points pilotes étant données un nombre suffisant et un bon positionnement des points pilotes. Dans le cas de la déformation graduelle, l'application peut réussir à condition que l'ensemble de fond soit assez grand. Cependant, cette paramétrisation augmente la non-linéarité inhérente de l'algorithme et peut entraîner des effets défavorables. Pour les deux cas, certains scénarios d'amélioration sont proposés et d'autres applications à des scénarios plus complexes sont conseillées.

Mots clés :

Calage historique, Assimilation de données, Réservoir, Filtre de Kalman d'ensemble (EnKF), Points pilotes, Déformation graduelle, Paramétrisation

Abstract

History-matching is an essential step in any reservoir study. It enables integration of data acquired after the production in the reservoir model building workflow. There exist several methods to perform such an act. Ensemble Kalman Filter (EnKF) is a sequential assimilation or history-matching method capable of integrating the measured data as soon as they are obtained. Several successful applications of the method are known in oceanographic, meteorological and petroleum applications. However, the performance of the method may deteriorate when the size of the ensemble is small or when dealing with non Gaussian distributions or highly nonlinear dynamical models.

This work is based on the EnKF application for History-matching purposes and is divided into two main sections.

First section deals with the application of the EnKF to several case studies in order to better understand the merits and shortcomings of the method. These case studies include two synthetic case studies (a simple one and a rather complex one), a Facies model and a real reservoir model. In most cases the method is successful in reproducing the measured data. The encountered problems, such as parameter overestimation for the rather complex synthetic case, are explained and possible solutions are proposed.

Second section deals with two newly proposed algorithms combining the EnKF with two parameterization methods: pilot point method and gradual deformation method. Both methods are capable of preserving second order statistical properties (mean and covariance). Their integration with the EnKF ensures preservation of the prior model properties during the assimilation process. Both developed algorithms are applied to the simple synthetic case study. For the pilot point method, the application was successful through an adequate number and proper positioning of pilot points. In case of the gradual deformation, the application can be successful provided the background ensemble is large enough. However, this parameterization increases the inherent nonlinearity in the algorithm and may result in unfavorable effects. For both cases, some improvement scenarios are proposed and further applications to more complex scenarios are recommended.

Key words:

History-matching, Data assimilation, Reservoir, Ensemble Kalman Filter (EnKF), Pilot points, Gradual deformation, Parameterization

Table des matières

Table de tableaux

Table des figures

1 Caractérisation de réservoir et calage d'historique

L'étude d'un gisement pétrolier est un travail multidisciplinaire faisant appel à diverses compétences en géosciences comme la géologie, la géophysique et l'ingénierie de réservoir. Le but est de construire un modèle le plus représentatif possible d'un réservoir en termes de prévision de la production, de quantification de l'incertitude et d'analyse des ressources pétrolières. Ceci a motivé le développement de workflows de simulations cherchant à contraindre des modèles de réservoir à toutes les données recueillies sur le terrain: données de diagraphie, mesures sur carottes, données sismiques, données de production, etc. Les étapes principales de ce workflow sont les suivantes (Figure 1) :

- Modélisation géologique

  Pour décrire la structure d'un réservoir pétrolier, les géologues s'appuient sur les informations fournies par les diagraphies, les carottes, les modèles dépositionnels, sédimentologiques et stratigraphiques existants, et les affleurements analogiques. Ces données sont dites "statiques" car elles ne varient pas dans le temps. Les géologues en déduisent des caractéristiques spatiales des propriétés pétrophysiques (porosité, perméabilité) généralement décrites par des statistiques d'ordre 2. Ceci permet d'utiliser des algorithmes géostatistiques pour générer des réalisations équiprobables des propriétés pétrophysiques qui peuplent un maillage couvrant le réservoir considéré: c'est le modèle géologique.

- Mise à l'échelle – Upscaling

  La simulation de l'écoulement des fluides dans un milieu poreux peut être très couteuse en temps de calcul si la grille considérée contient un grand nombre de mailles. On considère donc en général une deuxième grille plus grossière: les propriétés pétrophysiques y sont calculées à partir des valeurs du modèle géologique dans les mailles fines contenues dans chaque maille grossière. On parle alors de modèle réservoir grossier.

- Simulation d'écoulement

  L'écoulement des fluides dans un milieu poreux est décrit par un ensemble d'équations complexes, pour lesquelles une solution analytique est connue dans le cas d'un réservoir simplifié saturé en eau et huile (Cossé, 1993). Dans les autres cas, et notamment les cas réels, l'hétérogénéité du réservoir nécessite de résoudre ces équations numériquement, en

utilisant par exemple la méthode des différences finies (voir par exemple Cossé (1993) et Le Ravalec (2005)). On parle alors de simulateur d'écoulement.

- Calage d'historique

Avant mise en production d'un réservoir, seules les données statiques sont disponibles. Elles dépendent de façon linéaire des propriétés pétrophysiques et leur intégration dans le modèle réservoir se fait simplement. Dès que la production commence, le comportement dynamique du réservoir peut être observé aux puits par les outils mesurant pression et débit par exemple. Ces données sont dites "dynamiques" car elles varient dans le temps. On distingue en général deux types de données dynamiques. Le premier correspond aux propriétés associées à chaque maille du modèle réservoir telles que la pression et la saturation, appelées "variables dynamiques" dans la suite de ce document. Le second englobe les données de production mesurées aux puits comme la pression de fond (BHP), le débit de production d'huile (SOR) ou la percé d'eau (WCT), que l'on nommera valeurs mesurées, observations ou encore données de production.

L'intégration des données de production dans le modèle réservoir, appelée généralement "calage d'historique", est rendue difficile par le fait que ces données dépendent non-linéairement des propriétés du réservoir. Le principe est d'ajuster progressivement les paramètres décrivant le réservoir, tels que les propriétés pétrophysiques, les coefficients d'activité des aquifères, la composition des fluides, *etc.*, jusqu'à ce que le modèle obtenu reproduise les données de production observées. Ce processus peut être effectué de façon "assistée" grâce à une boucle d'optimisation. Les méthodes dites variationnelles reposent sur la minimisation d'une fonction objectif mesurant l'écart entre données simulées et données mesurées, en considérant simultanément toutes les observations disponibles. Pour intégrer de nouvelles mesures, le processus d'optimisation doit être relancé à partir du meilleur modèle existant, en comparant toutes les données disponibles aux réponses simulées. Une autre famille de méthodes, dites méthodes d'assimilation séquentielles, intègrent quant à elles les nouvelles mesures sans reconsidérer l'historique passé. Parmi elles, le filtre de Kalman d'ensemble (noté EnKF) a suscité un fort engouement ces dernières années en ingénierie de réservoir, permettant notamment de traiter les données provenant des systèmes de surveillance en continu. Sa mise en œuvre se révèle simple et indépendante du simulateur d'écoulement utilisé. En outre, elle fournit un ensemble de modèles calés, permettant d'estimer l'incertitude sur les prévisions de production.

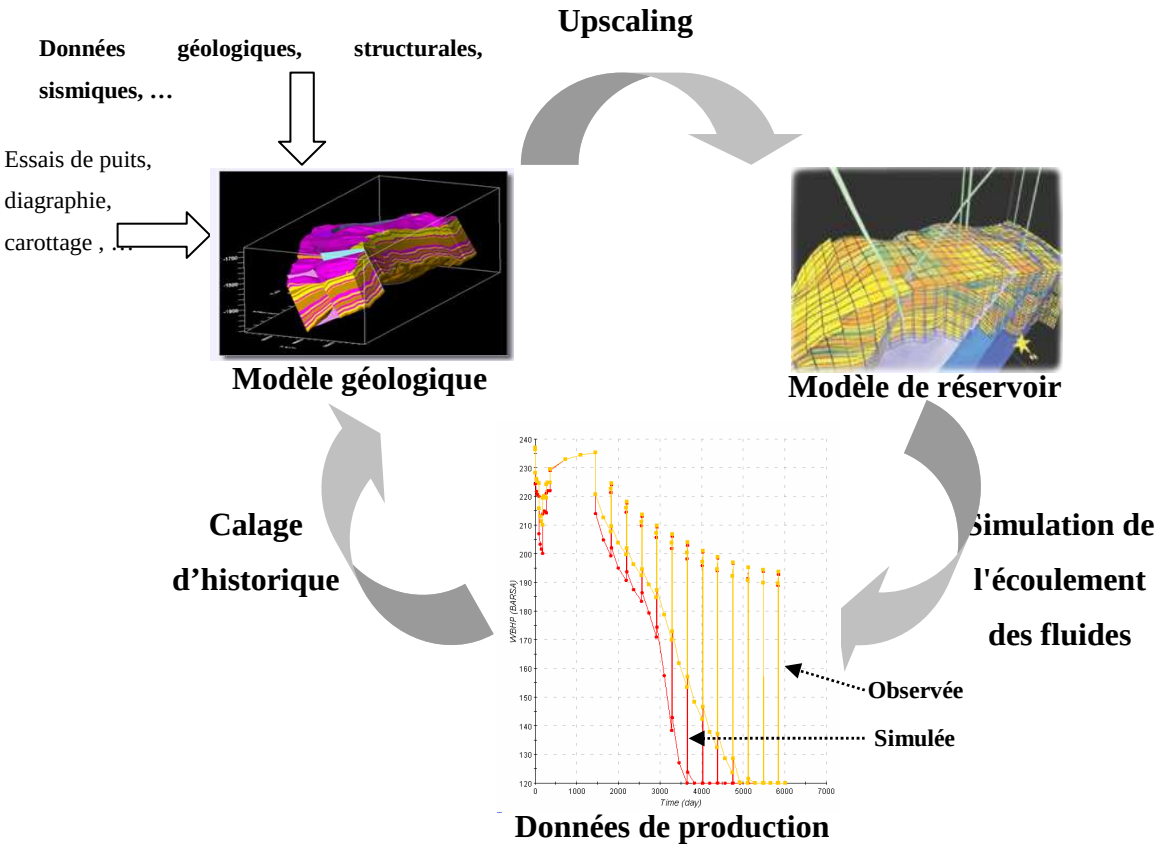


Figure 1 Workflow standard pour la construction de modèles de réservoir.

Le travail de thèse présenté ici vise à étudier la méthode des filtres de Kalman d'ensemble comme technique de calage d'historique dans le contexte pétrolier. En particulier, nous avons cherché à mettre en évidence les avantages et limites de cette méthode à travers plusieurs cas d'application, en termes de qualité du calage et des prévisions (section 3). De nouveaux algorithmes sont également proposés afin de préserver la variabilité spatiale des distributions des propriétés pétrophysiques au cours des assimilations. L'idée est de coupler la méthode EnKF avec deux techniques de paramétrisation existantes: la déformation graduelle et les points pilotes. Ces nouveaux algorithmes sont décrits dans la section 04.

2 Filtre de Kalman d'ensemble (EnKF)

Dans de nombreux domaines tels que l'écoulement des fluides en milieu poreux, l'un des problèmes majeurs consiste à identifier l'évolution d'un système au cours du temps en se basant sur une suite de mesures bruitées de ce système.

On définit l'état d'un système dynamique, ψ, par l'ensemble des variables nécessaires à sa description à un instant donné, appelées variables d'état. Afin d'analyser et de faire des prévisions sur l'évolution de ce système au cours du temps, au moins deux modèles sont nécessaires: le "modèle dynamique" qui décrit l'évolution de l'état du système au cours du temps, et le "modèle de mesure" qui lie les mesures bruitées à l'état du système (Arulampalam *et al.*, 2002).

2.1 Filtre de Kalman

Le filtre de Kalman (Kalman, 1960) est une méthode d'assimilation séquentielle de données permettant d'estimer l'état d'un système dynamique linéaire à partir de mesures bruitées. Il se déduit de la formule de Bayes, en cherchant soit une estimation du maximum de vraisemblance (Lorenc, 1986), soit une estimation du minimum de variance (Cohn, 1997). Le filtre de Kalman peut également être formulé comme un problème moindre carré récursif, pondéré par l'inverse de la matrice de covariance de l'erreur du modèle (Wunsch, 1996). Il s'appuie sur quatre hypothèses principales: le modèle dynamique doit être linéaire et constituer un processus de Markov du premier ordre; les variables d'état doivent suivre des distributions Gaussiennes; les erreurs sur le modèle et les erreurs de mesure doivent être indépendantes et non corrélées dans le temps.

La méthode du filtre de Kalman se décompose en deux étapes, appliquées sur chaque intervalle de temps entre deux acquisitions de mesures : l’ébauche (ou prévision) et la correction (ou analyse).

2.1.1 L'ébauche

Au cours du pas de temps k, le modèle dynamique f est appliqué pour faire évoluer l'état initial, ψ_{k-1}^{a}, au cours du pas de temps, fournissant ainsi l'ébauche (ou prévision) de l'état du système ψ_k^f à la fin du pas de temps:

$$\psi_k^f = f(\psi_{k-1}^a) + q_k \qquad 1.$$

Le terme q_k représente l'erreur du modèle dynamique au cours du pas de temps. L'incertitude sur l'état ψ_k^f est caractérisée par sa covariance $C_{\psi\psi}^{f,k}$. Si le modèle dynamique est linéaire, $C_{\psi\psi}^{f,k}$ peut être calculé directement (Evensen, 2007). Si le modèle est légèrement non linéaire, cette covariance peut être estimée par une expansion de Taylor de f : on parle alors de filtre de Kalman étendu (EKF). Pour un modèle dynamique fortement non linéaire, le filtre de Kalman d'ensemble (EnKF) apparaît plus approprié. Il repose sur l'application du filtre de Kalman à un ensemble d'états du système, utilisé pour estimer la covariance. Pour un nombre infini d'états, $C_{\psi\psi}^{f,k}$ peut être déterminée par (l'indice k a été enlevé pour plus de visibilité):

$$C_{\psi\psi}^{f,k} = \overline{(\psi^f - \psi^t)(\psi^f - \psi^t)^T} \qquad 2.$$

ψ^t représente l'état réel du système et la barre correspond à l'espérance. Dans les applications réelles, le nombre de réalisations disponibles N est limité et l'état réel du système n'est pas connu. On utilise alors l'estimation suivante:

$$C_{\psi\psi}^{f,k} \cong \overline{(\psi^f - \overline{\psi})(\psi^f - \overline{\psi})^T} \qquad 3.$$

où la barre générale désigne la moyenne sur l'ensemble des états et $\overline{\psi}$ l'état moyen calculé sur l'ensemble.

2.1.2 La correction

Lors de l'étape de correction (ou mise à jour), le modèle prédit ψ_k^f est contraint aux valeurs d_k mesurées à la fin du pas de temps k par :

$$\psi_k^a = \psi_k^f + K_k(d_k - H\psi_k^f) \qquad 4.$$

où

$$K_k = C_{\psi\psi}^{f,k} H^T (H C_{\psi\psi}^{f,k} H^T + C_{\varepsilon\varepsilon})^{-1} \qquad 5.$$

représente la matrice de gain et $C_{\varepsilon\varepsilon}$ la matrice de covariance des valeurs mesurées. Dans le cas du filtre de Kalman et du filtre de Kalman étendu, le calcul de la covariance $C_{\psi\psi}^{a,k}$ du modèle corrigé est simple. Dans le cas du filtre de Kalman d'ensemble, la formule (4) est appliquée à chaque modèle de l'ensemble de prévision. La covariance du nouvel ensemble obtenu est ensuite estimée par:

$$C_{\psi\psi}^{a,k} \cong \overline{(\psi^a - \overline{\psi})(\psi^a - \overline{\psi})^T} \qquad 6.$$

En outre, Burger *et al.* (1998) ont montré que les valeurs mesurées devaient être traitées comme des variables aléatoires afin de préserver la variabilité parmi les membres de l'ensemble au cours des assimilations. Ces derniers sont donc contraints à des valeurs mesurées perturbées, différentes d'un membre à l'autre.

2.2 Mise en œuvre de la méthode EnKF pour le calage d'historique

L'état d'un réservoir pétrolier, ψ, est caractérisé à chaque pas de temps par deux types de variables: des variables statiques telles que la porosité et la perméabilité, notées m par la suite, et des variables dynamiques comme la pression et la saturation, notées $g(m)$. La fonction g correspond à l'ensemble des équations décrivant l'écoulement en milieu poreux au cours du temps (modèle dynamique). Par ailleurs, durant la vie d'un réservoir, plusieurs données peuvent être mesurées aux puits comme la pression de fond (BHP), le débit de production d'huile (SOR) ou la percé d'eau (WCT).

Le vecteur d'état pour chaque membre de l'ensemble j au pas de temps k est donc:

$$\psi_{k,j} = \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \end{bmatrix} \qquad 7.$$

La relation entre valeurs mesurées, variables statiques et variables dynamiques n'étant pas linéaire, les réponses en production simulées $d_{sim}^{k,j}$ sont généralement ajoutées dans le vecteur d'état:

$$\psi_{k,j} = \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \\ d_{sim}^{k,j} \end{bmatrix} \qquad 8.$$

avec

$$d_{sim}^{k,j} = \begin{bmatrix} 0 & 0 & I \end{bmatrix} * \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \\ d_{sim}^{k,j} \end{bmatrix} = H_k * \psi_{k,j} \qquad 9.$$

où on a noté H_k l'opérateur de mesure et I la matrice identité. Les étapes d'ébauche et de correction sont ensuite appliquées à ce vecteur (équations 1 à 5).

2.3 Application de la méthode du filtre de Kalman d'ensemble dans l'industrie pétrolière

Dans l'industrie pétrolière, la méthode EnKF a été utilisée en premier lieu pour le calage en temps réel de données de forage (Lorentzen *et al.* ,2001). Ensuite, Nævdal *et al.* (2002) l'ont utilisé pour contraindre le champ de perméabilité d'un modèle à deux dimensions aux données de production, permettant ainsi de reproduire le modèle de perméabilité et d'obtenir de meilleures prévisions.

Les modèles en facies n'étant pas continus, le filtre de Kalman d'ensemble ne peut pas leur être appliqué directement. Dans Liu et Oliver (2005), les paramètres considérés sont donc les champs Gaussiens tronqués selon des seuils dans la méthode pluri-Gaussienne.

La méthode EnKF a également été appliquée au calage de données sismiques 4D. Ces dernières donnent l'évolution des attributs sismiques entre deux acquisitions, ce qui nécessite de combiner le filtre de Kalman d'ensemble et une de ses variantes dite EnKS (Ensemble Kalman Smoother). Skjervheim *et al.* (2005) ont appliqué cette méthode sur deux cas d'étude : un cas synthétique et un cas de mer du Nord. Pour le cas synthétique, l'assimilation simultanée des données de production et de sismique 4D a permis une meilleure estimation du champ de perméabilité. Pour le cas réel, le calage des données de production a été légèrement amélioré en considérant les deux types de données.

Dans Haugen *et al.* (2006) est présentée une application de la méthode EnKF à un champ de mer du Nord pour le calage de données de production. Des paramètres incertains déterministes sont inclus dans le vecteur d'état: transmissibilité des failles, profondeur du contact eau-huile et huile-gaz. Ce même type de paramètres est considéré dans Evensen *et al.* (2007) sur un autre champ de mer du Nord, assorti d'une discussion sur les avantages de la méthode EnKF par rapport à d'autres méthodes de calage d'historique.

2.4 Difficultés de la méthode EnKF

Bien que la plupart des applications de la méthode EnKF au calage d'historique soient considérées comme des succès, plusieurs problèmes peuvent affecter ses performances. Ils sont liés notamment à la taille finie de l'ensemble, à l'hypothèse de distribution Gaussienne des variables d'état, ou encore à la dépendance fortement non linéaire entre variables d'état et mesures.

La taille finie de l'ensemble peut générer des artefacts dans la matrice de covariance, comme une corrélation forte entre deux points très éloignés. Une solution consiste à utiliser des méthodes dites de localisation, en calculant par exemple un voisinage autour de chaque point hors duquel la covariance est mise à zéro (Houtekamer and Mitchell, 2001). D'autres méthodes sont proposées dans Furrer et Bengstone (2007) ou Arroyo-Negrete *et al.* (2007) par exemple.

Lorsque les variables d'état n'ont pas une distribution Gaussienne, une paramétrisation peut être utilisée pour les remplacer par d'autres variables suivant une loi Gaussienne. Parmi ces méthodes, on peut citer la transformation lognormale appliquée par exemple dans les travaux de Bertino *et al.* (2003) et Gu et Oliver (2006).

Enfin, des approches itératives peuvent être utilisées lorsque la dépendance entre les variables d'état et les valeurs mesurées est fortement non linéaire (Reynolds *et al.* (2006), Gu et Oliver (2007)).

3 Cas d'étude

L'objectif de ce travail est d'étudier les performances de la méthode EnKF pour le calage d'historique à travers différents cas d'étude synthétiques.

- Cas 2D simple (section 3.1)

  L'objectif de ce cas est d'améliorer l'estimation d'un champ de perméabilité par l'assimilation de données de production. Le cas d'étude est assez simple et ne présente pas de changement abrupt dans l'écoulement des fluides. L'effet de plusieurs paramètres est étudié : taille de l'ensemble, pas de temps d'assimilation, incertitude sur les valeurs mesurées et choix de l'ensemble initial.
- Cas 3D issu d'un cas réaliste: variante du cas PUNQ-S3 (section 3.2)

  Le but est d'examiner les performances de la méthode EnKF sur un cas plus complexe, avec changements abrupts dans l'écoulement des fluides liés à des variations des débits de production cible.
- Cas 2D avec modèle en faciès (section 3.3)

  On considère un réservoir synthétique dans lequel plusieurs faciès sont présents. Les modèles en faciès ne peuvent pas être directement corrigés par la méthode EnKF en raison de leur caractère discontinu. Ici, les variables d'état considérées sont donc les valeurs du champ Gaussien utilisé pour générer la distribution en faciès par la méthode de la Gaussienne tronquée.

3.1 Cas 2D simple

On considère ici une grille 2D contenant 50×50 mailles de dimensions 10 m × 10 m. La méthode FFT-MA (Fast-Fourier Transform Moving Average, Le Ravalec *et al* (2005)) est utilisée pour générer un champ de perméabilité de référence sur cette grille (Figure 2). Après transformation logarithmique, la propriété $\ln(kh)$ est simulée avec une moyenne de 5, une variance de 2.5, et un variogramme sphérique anisotrope, la dimension principale d'anisotropie faisant un angle de 30 degrés par rapport à l'axe X. Les longueurs de corrélation sont de 200 et 40 m respectivement.

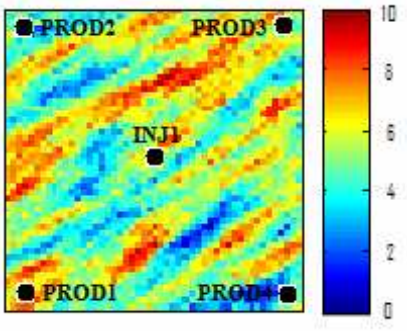


Figure 2 Champ ln(kh) de référence

Cinq puits verticaux sont présents dans le réservoir : un injecteur d'eau (INJ) au centre et quatre producteurs dans les coins (PROD1 à PROD4, voir Figure 2). Le débit d'huile cible imposé aux producteurs est de 40 m^3/jour. Le débit d'injection d'eau est initialement de 40 m^3/jour, puis passe à 50 m^3/jour et 60 m^3/jour après 90 et 150 jours de production, respectivement.

L'objectif de cette expérience est de régénérer le champ de référence $\ln(k)$ par l'assimilation des données de production. A ce stade, nous supposons donc connues les données de production simulées avec le champ de référence, mais pas le champ lui-même. Ces données sont la pression de fond (BHP), le débit d'huile (SOR) et la percée d'eau (WCT) représentées sur la Figure 3 par les courbes rouges. Puis nous allons utiliser la méthode EnKF pour identifier un ensemble de réalisations de ln(k) conditionnées à ces données de production, en effectuant une assimilation à la fin de chaque mois pendant 516 jours (17 mois).

La distribution de l'erreur sur les mesures est supposée Gaussienne avec un écart-type égal à 5% des valeurs mesurées.

3.1.1 Calage d'historique avec la méthode EnKF

Un ensemble initial de 50 réalisations de la perméabilité est généré avec l'algorithme FFT-MA, en utilisant les mêmes propriétés statistiques que pour le modèle de référence. Les autres paramètres (porosité, perméabilité relative, pression initiale, *etc.*) sont supposés connus.

Le vecteur d'état pour chaque membre j de l'ensemble contient les variables suivantes :

- Le logarithme de la perméabilité dans chaque maille (variable statique) : $\ln k_{j,i}, i = 1,...,2500.$
- La pression dans chaque maille (variable dynamique) : $P_{j,i}, i = 1,...,2500.$

- La saturation en eau dans chaque maille (variable dynamique) : $Sw_{j,i}, i = 1,...,2500.$
- La pression de fond aux puits : $BHP_{j,k}, k = 1,...,5.$
- Le débit de production d'huile aux producteurs : $SOR_{j,k}, k = 1,...,4.$
- La percée d'eau aux producteurs : $WCT_{j,k}, k = 1,...,4.$

Il est donc défini par :

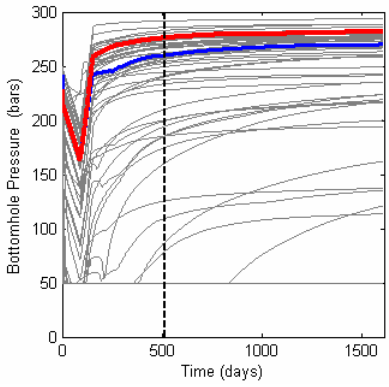
$$\psi_j = [\ln k_{j,1}, \ln k_{j,2}, ..., \ln k_{j,2500}, P_{j,1}, P_{j,2}, ..., P_{j,2500}, Sw_{j,1}, Sw_{j,2}, ..., Sw_{j,2500}, \quad 10.$$
$$BHP_{j,1}, ..., BHP_{j,5}, SOR_{j,1}, ..., SOR_{j,4}, WCT_{j,1}, ..., WCT_{j,4}]$$

La période d'assimilation est prolongée jusqu'à 1612 jours afin d'évaluer la qualité de prévision des modèles de perméabilité obtenus à l'issue du calage.

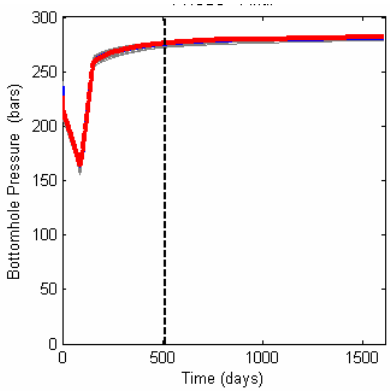
3.1.2 Résultats du calage d'historique

Nous comparons ici la performance des modèles initiaux et finaux pour le puits PROD3 (Figure 3). Celui-ci est situé dans une région fortement perméable qui s'étend jusqu'à l'injecteur. Le temps de percée d'eau le plus court y est observé.

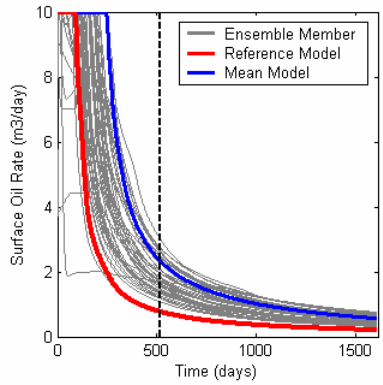
On peut voir que les données simulées avec les modèles initiaux présentent plus de variabilité que celles obtenues avec les modèles corrigés. Les Figure 3 (a)-(c) montrent que, pour certains modèles initiaux, les valeurs simulées sont très éloignées des valeurs de référence. Après assimilation, elles en sont toutes très proches (Figure 3 (d)-(f)), dans la phase de calage comme dans la période de prévision.



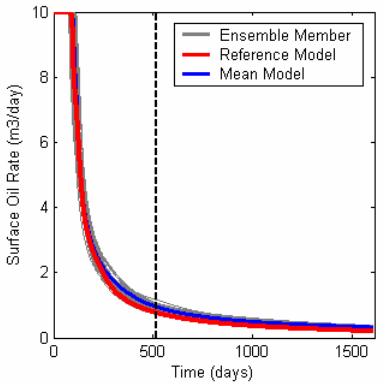
(a) Pression de fond – Ensemble initial



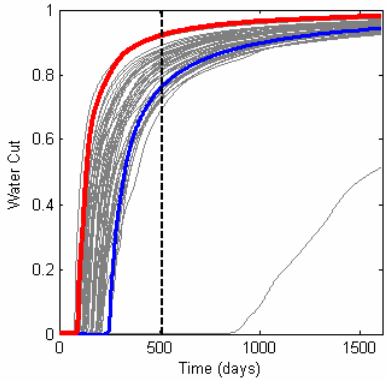
(d) Pression de fond – Ensemble final



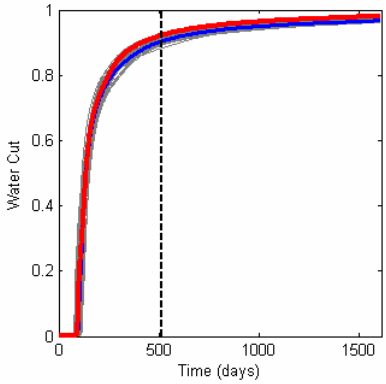
(b) Débit d'huile – Ensemble initial



(e) Débit d'huile – Ensemble final



(c) Percée d'eau – Ensemble initial



(f) Percée d'eau – Ensemble final

Figure 3 Données de production simulées au puits PROD3 avec les 50 réalisations de l'ensemble initial (première colonne) et de l'ensemble final (deuxième colonne). La ligne verticale à 516 jours indique la limite entre la période de calage (0-516 jours) et la période de prévision (517-1612 jours). Les courbes grises correspondent aux membres de l'ensemble et la courbe bleue au modèle moyen. Les valeurs de référence sont tracées en rouge.

L'assimilation de donnée nous fournit donc un ensemble de modèles corrigés de perméabilité pour lesquels les réponses en production simulées sont beaucoup plus proches des valeurs de référence que celles de l'ensemble initial.

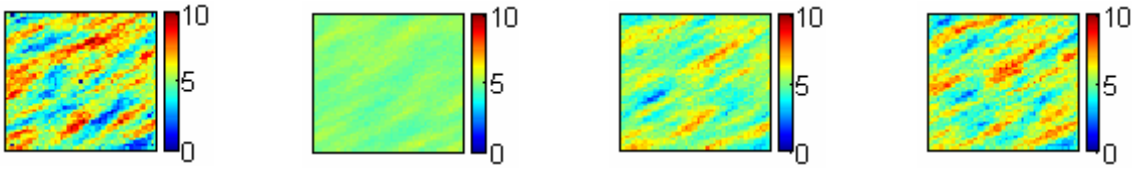
La qualité du calage pour les données de production peut être évaluée par l'erreur quadratique moyenne $RMS_{prod,1}$, qui mesure l'écart entre données mesurées et valeurs simulées avec chaque membre de l'ensemble:

$$RMS_{prod,1} = \sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}\frac{1}{N_t}\sum_{j=1}^{N_t}\frac{1}{N_{data}^{t_j}}\sum_{k=1}^{N_{data}^{t_j}}(d_{obs}^{k}(t_j)-d_i^{k}(t_j))^2} \qquad 11.$$

N_e est la taille de l'ensemble, N_t le nombre d'assimilations et $N_{data}^{t_l}$ le nombre de données à assimiler au pas de temps t_j (en général, celui-ci est donné par le nombre de puits où des données sont observées). i est l'indice du membre de l'ensemble, j l'indice du pas de temps

et k l'indice des données de production. $d_{obs}^{k}(t_j)$ représente la valeur mesurée au pas de temps k et $d_i^k(t_j)$ la valeur simulée correspondante. L'erreur $RMS_{prod,1}$ calculée pour l'ensemble initial et l'ensemble final montre ici une réduction de 8, 23 et 24% pour la pression, le débit d'huile et la percée d'eau respectivement.

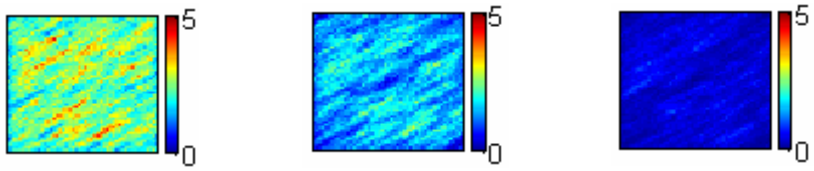
Un autre aspect important est la capacité de la méthode à fournir des modèles pétrophysiques proches du champ de référence. L'évolution du modèle de perméabilité moyen de l'ensemble au cours des assimilations est présentée sur la Figure 4. Le modèle moyen initial est uniforme. Après la quatrième assimilation, on voit que des hétérogénéités commencent à apparaître. Ce pas de temps correspond à la première percée d'eau aux producteurs, au puits PROD3. Le modèle moyen final reproduit la plupart des principales caractéristiques du modèle de référence. Toutefois, il apparaît plus lisse et présente peu d'hétérogénéités aux petites échelles.



(a) Modèle de référence **(b) Modèle initial** **(f) 4eme assimilation** **(s) 17eme assimilation**

Figure 4 Évolution des modèles au cours des assimilations : (a) Modèle de référence, (b) Modèle moyen initial, (c) Modèle moyen après la 4eme assimilation et (d) Modèle moyen après la 17eme assimilation.

La variance de l'ensemble à chaque pas de temps est présentée sur la Figure 5. Du fait des assimilations, l'incertitude sur le modèle, c'est-à-dire la variance de l'ensemble, diminue, en particulier autour des puits. Ici, la réduction est significative. Ce phénomène est une conséquence naturelle de la procédure d'assimilation, mais une trop forte diminution doit être évitée.



(a) Ensemble Initial **(b) 4eme assimilation** **(c) 17eme assimilation**

Figure 5 Évolution de la variance : (a) Ensemble initial, (b) Après la 4eme assimilation et (c) Après la 17eme assimilation.

3.1.3 Études de sensibilité

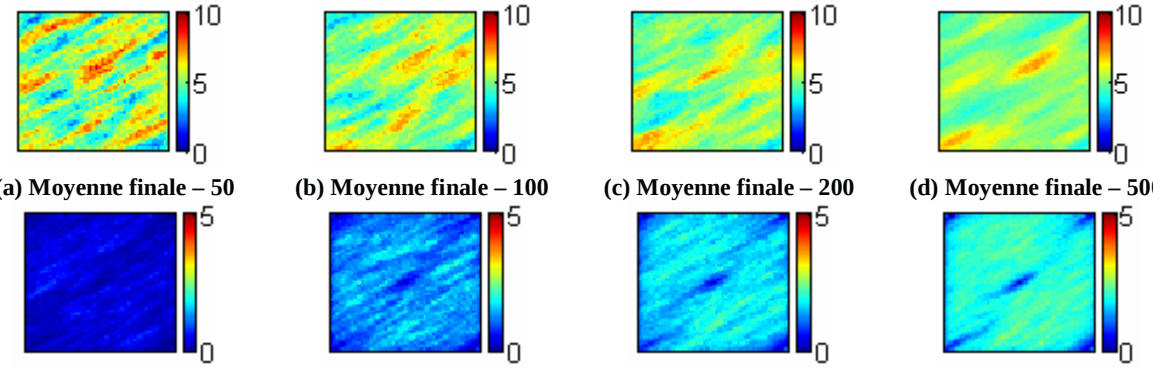
Plusieurs paramètres peuvent affecter la performance de la méthode EnKF, comme la taille de l'ensemble, le pas de temps d'assimilation, le niveau de l'incertitude ou le choix de

l'ensemble initial. Dans cette section, nous présentons une série de tests visant à étudier le comportement de la méthode pour différentes configurations de ces paramètres.

3.1.3.1 Effet de la taille de l'ensemble

La méthode EnKF estime la matrice de covariance des variables d'état en utilisant un ensemble de réalisations pour ces variables. En augmentant la taille de l'ensemble, N, l'erreur sur cette approximation diminue selon $1/\sqrt{N}$, ce qui devrait donc permettre de limiter les corrélations factices dans la matrice de covariance. Cependant, augmenter la taille de l'ensemble nécessite également plus de temps de calcul pour réaliser les simulations d'écoulement supplémentaires. Il s'agit donc de trouver un bon compromis entre la précision obtenue sur la matrice de covariance et le temps de calcul. Le choix de la taille de l'ensemble dépend de l'application considérée, mais un ensemble de taille 100 apparaît généralement consensuel dans la littérature. Dans cette section, nous considérons le même cas d'étude que précédemment, en faisant varier la taille de l'ensemble à 100, 200 et 500. Le niveau d'incertitude sur les valeurs mesurées et le pas d'assimilation sont inchangés.

Comme nous pouvons le voir sur la Figure 6 (a)-(d), le logarithme de la perméabilité moyen de l'ensemble final présente quelques caractéristiques spatiales similaires à celles du modèle de référence. En augmentant la taille de l'ensemble, cette moyenne apparaît plus lisse. La Figure 6 (e)-(h) présente la variance des ensembles finaux. On peut voir que cette dernière a diminué, notamment autour des puits. Toutefois, elle apparaît très faible pour des ensembles de petite taille, 50 notamment.



(a) Moyenne finale – 50 **(b) Moyenne finale – 100** **(c) Moyenne finale – 200** **(d) Moyenne finale – 500**

(e) Variance finale – 50 **(f) Variance finale – 100** **(g) Variance finale – 200** **(h) Variance finale – 500**

Figure 6 Moyenne et variance des modèles de ln(kh) finaux, pour des ensembles de taille 50, 100, 200 et 500.

En ce qui concerne les données de production, les valeurs simulées avec les ensemble finaux permettent de régénérer le modèle de référence mieux que les ensemble initiaux, et ce quelle que soit la taille de l'ensemble. De plus, l'écart de données de production après la période du

calage d'historique augment avec un ensemble plus grand. Ce fait est une conséquence directe de préservation de la variation dans les ensembles finaux. Les améliorations obtenues en utilisant un ensemble de taille 200 ou 500 plutôt qu'un ensemble de taille 100 sont moins importantes que lorsque l'on passe de 50 à 100. Un ensemble de taille 100 semble le plus approprié pour ce cas.

3.1.3.2 Effet du pas de temps d'assimilation

L'assimilation des données de production apparaît nécessaire lors de forts changements dans l'écoulement des fluides. Le pas de temps d'assimilation ne doit pas être trop court (pour éviter que les données ne soient corrélées et redondantes) ni trop long afin de bien capturer les non linéarités de l'écoulement. Trois expériences ont été effectuées pour déterminer l'effet de ce paramètre, en conservant les autres paramètres inchangés :

- 1^{ere} expérience : deux assimilations par mois.
- 2^{eme} expérience : une assimilation par mois (configuration utilisée précédemment).
- 3^{eme} expérience : une assimilation tous les deux mois.

En réduisant le pas d'assimilation, on constate que, globalement, le RMS des données de production diminue.

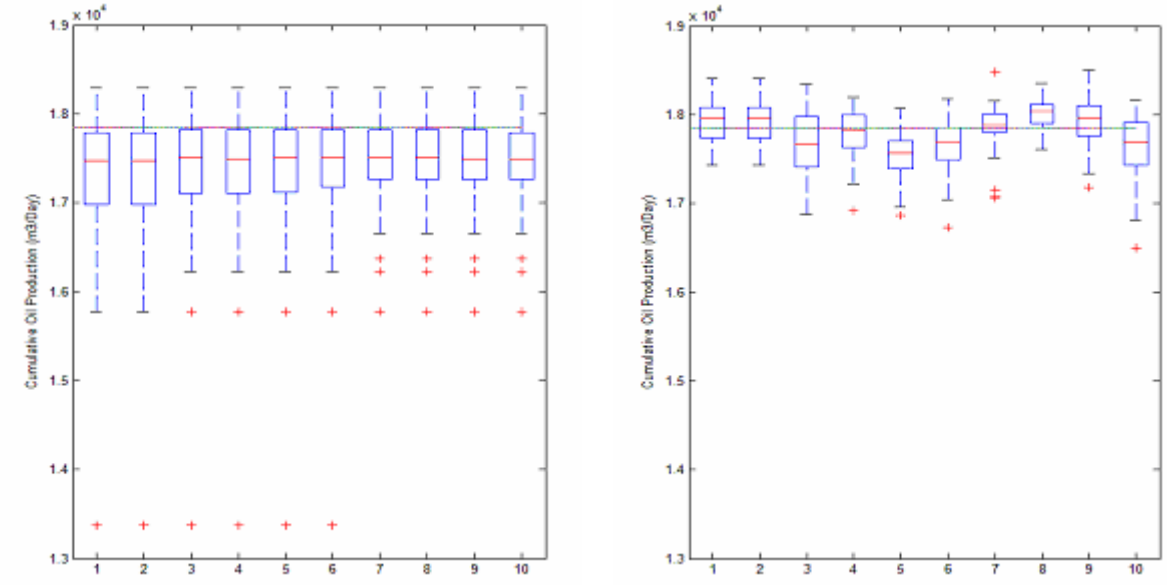
3.1.3.3 Effet de l'incertitude sur les valeurs mesurées

L'incertitude sur les valeurs mesurées, c'est-à-dire l'écart type du bruit sur les mesures, dépend de l'incertitude des outils de mesure, de leur positionnement, etc. Ainsi, elle varie d'un cas à l'autre et doit être fixée au début de la procédure d'assimilation. Ce niveau d'incertitude est repris ici pour générer l'ensemble des mesures perturbées. Le cas étant synthétique, nous avons considéré trois niveaux d'incertitude : 1%, 5% et 10% de chaque valeur mesurée. Les autres paramètres d'assimilation sont les mêmes que précédemment.
Les résultats obtenus suggèrent qu'une incertitude trop faible peut poser des problèmes en conduisant à une variance très faible de l'ensemble. Si ce problème est évité, le choix de l'incertitude sur les valeurs mesurées est de moindre importance.

3.1.3.4 Effet de l'ensemble initial

L'ensemble initial doit être représentatif de l'espace des modèles a priori. De plus, les modèles corrigés par la méthode EnKF deviennent dépendants, et un seul ensemble final ne suffit pas pour faire des prévisions fiables (Thulin *et al.*, 2008). On cherche ici à comparer les

résultats obtenus avec différents ensembles initiaux de même taille pour la prévision de la production cumulée d'huile.



(a) Ensembles initiaux **(b) Ensembles finaux**

Figure 7 Box plot de la production cumulée d'huile pour 10 ensembles (a) ensembles initiaux, (b) ensemble finaux.

A cette fin, le même scénario d'assimilation est appliqué à dix ensembles initiaux de taille 50. Les valeurs de la production cumulée d'huile simulées à la fin de la période de prévision avec les ensembles initiaux et finaux sont comparées sur la Figure 7. La ligne horizontale indique la valeur de référence. Elle apparaît donc dans la fourchette de valeurs simulées avec les ensembles initiaux. Après assimilation, les prédictions varient encore d'un ensemble à l'autre, même si les différences sont réduites. La valeur de référence est située entre les quartiles supérieur et inférieur pour la plupart des ensembles. Ainsi, nos résultats suggèrent que les prévisions sont plus crédibles lorsque l'expérience est répétée pour plusieurs ensembles de même taille.

3.2 Cas 3D issu d'un cas réaliste : variante du cas PUNQ-S3

Un cas d'étude plus complexe que celui considéré dans la section précédente est nécessaire pour examiner les performances de la méthode. On considère donc ici une variante du cas d'étude PUNQ-S3.

3.2.1 Modèle de réservoir et propriétés pétrophysiques

Le cas d'étude PUNQ-S3 est un modèle standard de petite taille défini à partir d'un champ réel opéré par Elf Exploration et Production pour réaliser des benchmarks. Une description complète de ce cas est disponible sur le site internet du projet PUNQ et dans Floris *et al*. (2001).

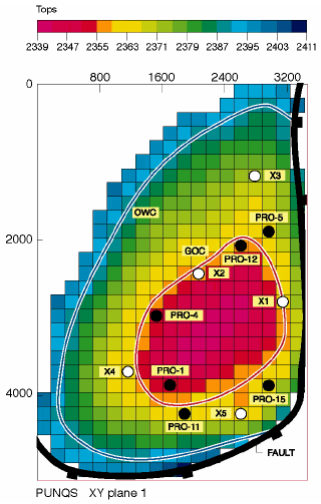


Figure 8 Toit du réservoir et positionnement des puits (issu du site internet du projet PUNQ)

Le modèle réservoir est construit sur une grille de taille 19×28×5. Les mailles ont pour dimension 180 m dans les directions X et Y, leur épaisseur étant variable. 1761 d'entre elles sont actives. La production est assurée par 6 puits : PRO-1, PRO-4, PRO-5, PRO-11, PRO-12 et PRO-15 (Figure 8). La distribution de la porosité et de la perméabilité horizontale est générée aléatoirement avec les caractéristiques données dans le Tableau 1, les valeurs de référence étant celles présentées sur les Figure 9 et Figure 10. La perméabilité verticale est égale à la perméabilité horizontale. On cherche ici à générer des modèles de porosité et perméabilité proches des valeurs de référence par l'assimilation des données de production.

Tableau 1 Caractéristiques des propriétés pétrophysiques (porosité et log de perméabilité) dans chaque couche du réservoir

Couche	1	2	3	4	5
Moyenne de la porosité	0.1722	0.0802	0.1677	0.1615	0.1892
Variance de la porosité	0.0078	0.0004	0.0050	0.0006	0.0049
Moyenne de ln (kh)	2.18	1.41	2.24	2.47	2.49
Variance de ln (kh)	3.14	0.74	3.26	5.64	3.72
Longueur de corrélation (m)	3500	750	6000	1500	3750
Ratio d'anisotropie	0.286	1.0	0.25	0.50	0.333
Azimut (degré)	30	0	45	-30	60

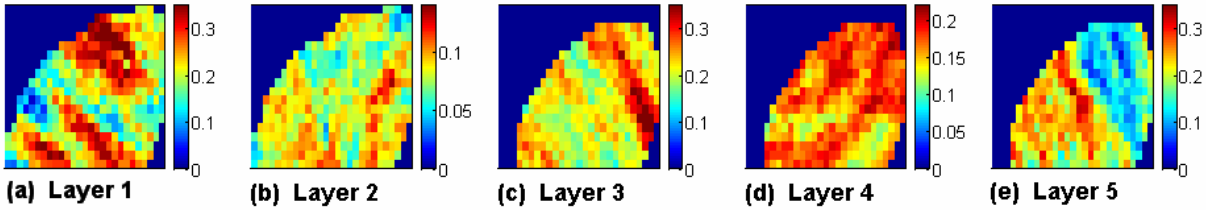


Figure 9 Porosité de référence dans les cinq couches.

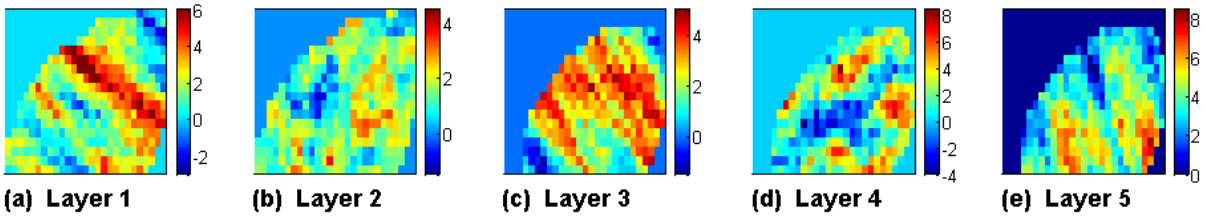


Figure 10 Ln(kh) de référence dans les cinq couches.

3.2.2 Schéma de production

Contrairement au cas PUNQ-S3, le réservoir ne contient ici que de l'eau et de l'huile (le gaz n'est pas considéré). Le schéma de production est le suivant :

- Une année d'essais de puits comprenant quatre périodes de production de trois mois pendant lesquelles des débits cibles de production d'huile de 100, 200, 100 puis 50 m^3/jour sont imposés aux puits.
- Trois années de fermeture des puits ;
- Quatre années de production, pendant lesquelles un débit cible de production d'huile de 100 m^3/jour est imposé, avec fermeture des puits pendant les deux dernières semaines de chaque année.

Ces 8 années constituent la période d'assimilation, prolongée par une période de prévision de 8.5 années. 13 pas d'assimilation sont considérés, qui correspondent aux changements des débits de production.

Les incertitudes sur les valeurs mesurées sont de 2 bars pour la pression, 2 m^3/jour pour les débits et 0.02 pour la percée d'eau.

3.2.3 Procédure d'assimilation

La taille de l'ensemble est fixée à 50. Étant donné que le modèle contient 1761 mailles actives, les variables d'état sont les 1761 valeurs correspondantes de porosité, perméabilité

(variables statiques), pression et saturation en eau (variables dynamiques) dans ces mailles. Le vecteur d'état pour le membre j et le pas de temps k est donc défini par :

$$\psi_{k,j} = [trans(\varphi_{k,j,1}),...,trans(\varphi_{k,j,N_a}),trans(kh_{k,j,1}),...,trans(kh_{k,j,N_a}),$$
$$P_{k,j,1},...,P_{k,j,N_a},S_{wk,j,1},...,S_{wk,j,N_a},d_{k,j,1},...,d_{k,j,N_d}] \quad (12.)$$

N_a est le nombre de mailles actives dans le modèle et N_d le nombre total de données de production. $\varphi_{k,j,i}$, $kh_{k,j,i}$, $P_{k,j,i}$ et $Sw_{k,j,i}$ représentent la porosité, perméabilité, pression et saturation en eau dans la maille i, et $d_{k,j,i}$ les données de production simulées. La fonction $trans$ définit la transformation appliquée aux variables statiques pour s'assurer que leur valeur soit encore physique après correction par le filtre :

1. Pour la porosité : $trans(\varphi) = \log(\frac{\varphi}{1-\varphi})$
2. Pour la perméabilité : $trans(kh) = \log(kh)$

3.2.4 Résultats du calage d'historique

Nous comparons ici la performance des modèles initiaux et finaux pour le puits PRO-5 (Figure 11). Comme pour le cas précédent, on peut voir que les valeurs simulées avec les membres de l'ensemble initial présentent une forte variation autour des valeurs de référence, certaines étant très éloignées de ces valeurs. Après assimilation, les valeurs simulées sont plus proches des données mesurées. Le modèle pétrophysique moyen calculé sur l'ensemble final reproduit mieux les valeurs de référence que celui calculé sur l'ensemble initial. L'assimilation permet également de beaucoup mieux reproduire les percées d'eau, même si le résultat obtenu n'est pas encore acceptable. De façon générale, les données de percée d'eau sont plus difficiles à caler du fait de leur dépendance très fortement non-linéaire avec les variables statiques. Les prévisions obtenues pour la pression et le débit d'huile apparaissent raisonnables, tandis que celles pour la percée d'eau sont peu fiables suite au mauvais calage de ces données.

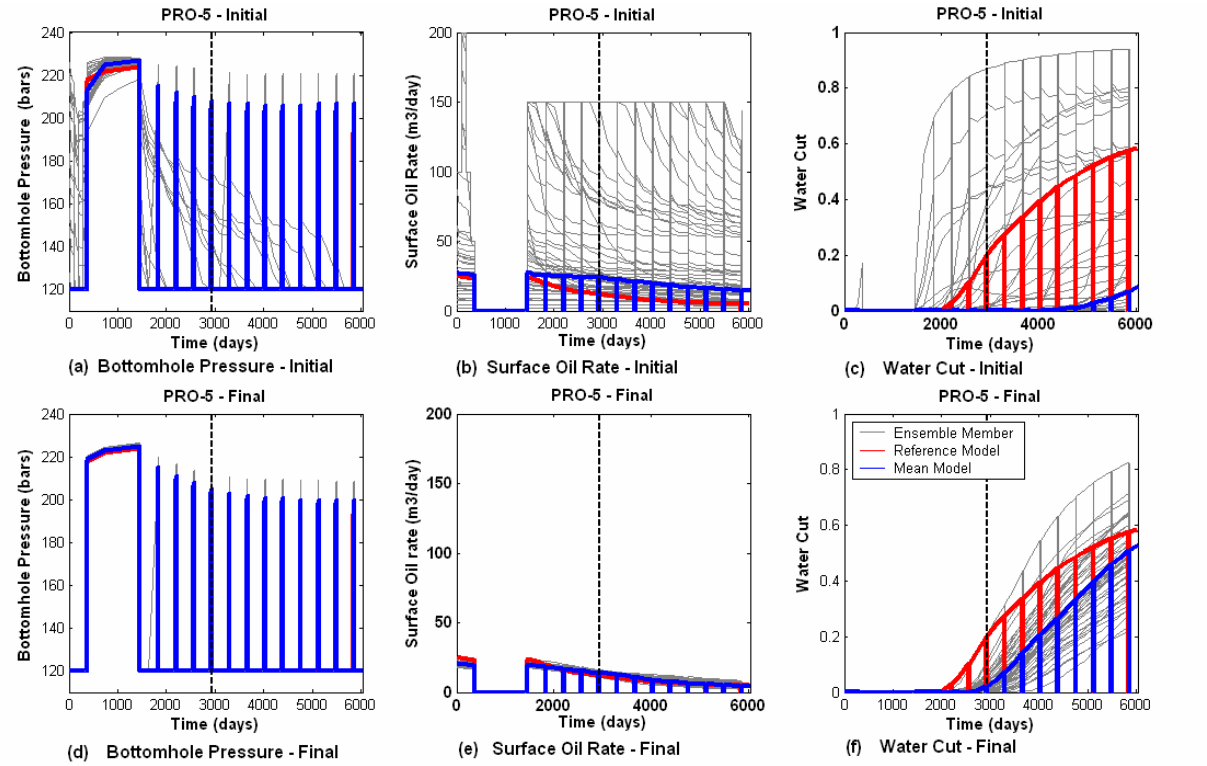


Figure 11 Données de production simulées au puits PRO-5 avec les 50 réalisations de l'ensemble initial (première ligne) et final (deuxième ligne). La ligne verticale à 2936 jours indique la limite entre la période de calage (0-2936 jours) et la période de prévision (2937-6025 jours). Les courbes grises correspondent aux membres de l'ensemble et la courbe bleue au modèle moyen. Les valeurs de référence sont tracées en rouge.

Nous avons réitéré le processus d'assimilation à partir de quatre autres ensembles initiaux de taille 50. Ceci nous a permis d'observer que le RMS sur les données de production variait significativement d'un ensemble à l'autre, surtout pour la percée d'eau. Cependant, ces différences s'estompent pour des ensembles de taille 100.

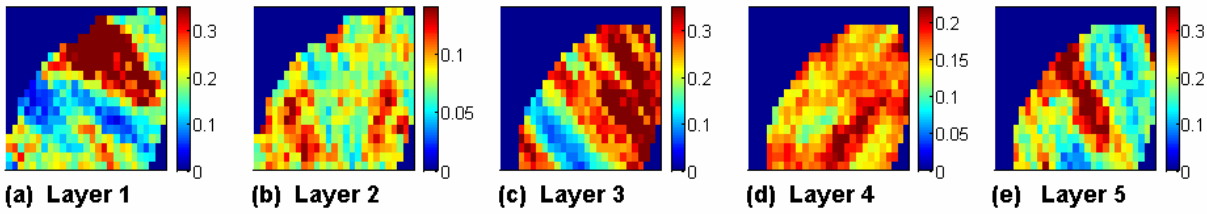


Figure 12 Porosité moyenne de l'ensemble final dans les 5 couches.

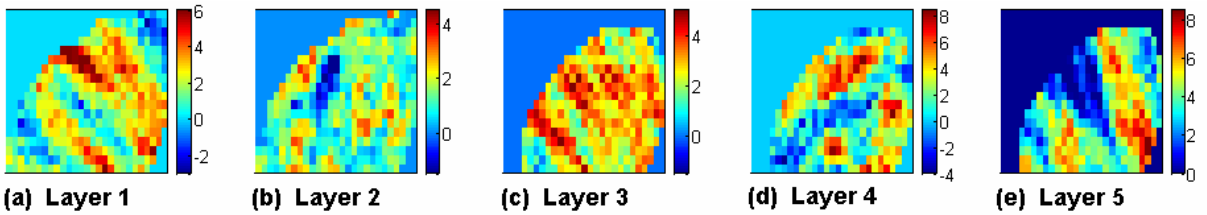


Figure 13 Perméabilité moyenne (ln kh) de l'ensemble final dans les 5 couches.

Les modèles moyens de porosité et perméabilité (ln(kh)) obtenus après 13 assimilations sont présentés sur les Figure 12 et Figure 13. On y retrouve les principales caractéristiques des modèles de référence. Par ailleurs, la porosité est surestimée dans les couches 1, 3 et 5. Selon Devegowda *et al.* (2007), cette surestimation est due aux corrélations fausses dans la matrice de covariance, induites par la taille finie de l'ensemble. Ce problème peut être évité par des méthodes de localisation.

3.2.5 Études de sensibilité

Nous avons effectué une série d'études de sensibilité sur les mêmes paramètres que pour le cas synthétique 2D, avec des conclusions très proches.

- En augmentant la taille de l'ensemble, les données de production sont mieux calées. Ceci est le plus visible quand la taille de l'ensemble passe de 50 à 100, une augmentation à 200 ou 500 ayant moins d'effet. Ici, il semble qu'une taille de 100 à 200 soit appropriée pour le calage avec la méthode EnKF.
- Le niveau d'incertitude sur les valeurs mesurées n'est pas un paramètre influent pour l'assimilation des données.
- En diminuant le pas de temps d'assimilation, les changements significatifs dans l'écoulement des fluides sont mieux capturés. Pour une taille d'ensemble donnée, une diminution du pas de temps améliore le calage des données de production. Ceci est d'autant plus visible lorsque l'on augmente simultanément la taille de l'ensemble.

3.3 Cas 2D avec modèle en faciès

Les modèles en faciès sont utilisés pour décrire différents types de roche et milieux de dépôts dans les réservoirs. Dans un tel modèle, il existe donc un nombre fini de faciès avec des propriétés pétrophysiques spécifiques. Ces modèles n'étant pas continus, les contraindre aux données dynamiques avec la méthode EnKF ne peut se faire directement en les considérant dans le vecteur d'état. Ici, nous utilisons la méthode de la Gaussienne tronquée pour générer des distributions en faciès, ce qui nous a conduit à considérer dans le vecteur d'état le champ Gaussien tronqué selon les proportions de chaque faciès.

3.3.1 Méthode de la Gaussienne tronquée

Supposons que le réservoir contienne m faciès. Pour simuler ces faciès, on peut par exemple appliquer $m-1$ seuils de troncature à une fonction aléatoire Gaussienne centrée réduite et attribuer chacune des m zones obtenues à un des faciès. Ceci est à la base de la méthode de la Gaussienne tronquée (Figure 14).

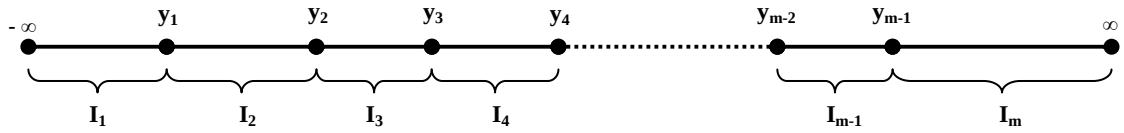


Figure 14 Méthode de la Gaussienne tronquée. y_i est le seuil correspondant au faciès I_i.

La fonction aléatoire Gaussienne $Y(x)$ prend ses valeurs dans $\Re$. Les fonctions indicatrices $I_i(x), i=1,...,m$ représentent chacun des m faciès définis par les seuils $y_i . i=1,...,m$:

$$I_i(x) = 1_{y_{i-1} \le Y(x) < y_i} \qquad -\infty = y_0 < y_1 < ... < y_{i-1} < y_i < ... < y_m = +\infty \qquad 13.$$

Ces seuils sont choisis de manière à reproduire les proportions de chaque facies p_i :

$$y_i = G^{-1}\left(\sum_{j=1}^{i} p_j\right) \ , \ i=1,...,m-1 \qquad 14.$$

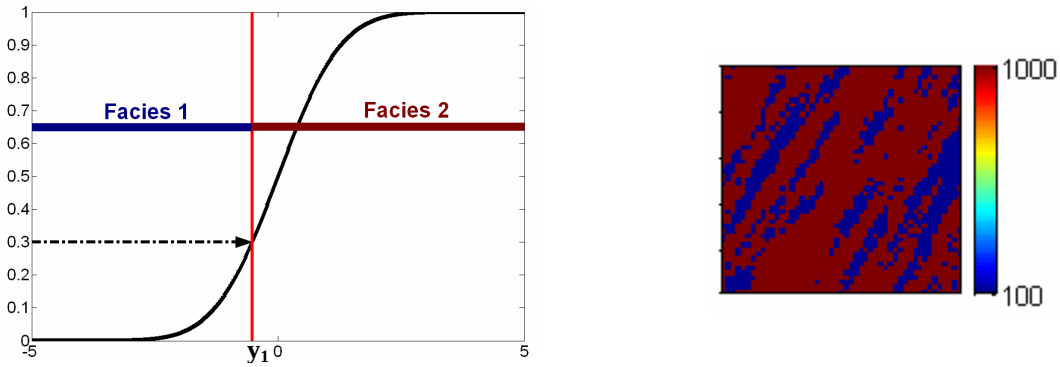
Dans cette formule, G est la fonction de répartition d'une variable Gaussienne centrée réduite.

La méthode de la Gaussienne tronquée est appropriée pour les modèles en faciès simples. Pour les cas plus complexes, son extension, la méthode pluri-Gaussienne, peut être plus appropriée. Pour l'ajustement des modèles en facies avec la méthode EnKF, la fonction $Y(x)$ est considérée comme la variable statique. Les seuils de troncature sont ensuite appliqués pour déterminer la distribution en faciès.

3.3.2 Modèle de réservoir

Le modèle réservoir est construit sur la même grille que le premier cas d'étude, de taille 50×50. Deux types de facies sont considérés : l'un possède une faible perméabilité de 100 mD et une proportion de 30%, tandis que le deuxième est très perméable (perméabilité de 1000 mD). La porosité est uniforme et vaut 0,18. Le champ Gaussien auquel est appliqué la troncature est généré par l'algorithme FFT-MA, en considérant un variogramme sphérique et

une direction principale d'anisotropie définie par un angle de 60 degrés par rapport à l'axe X. Les longueurs de corrélation sont 300 et 40 m respectivement.



(a) Plan de seuillage pour les faciès **(b) Perméabilité de référence**

Figure 15 (a) Plan de seuillage pour les faciès (b) Perméabilité de référence : le bleu correspond au faciès 1 et le rouge au faciès 2.

Le principe de troncature décrit sur la Figure 15 (a) est appliqué pour attribuer les valeurs de faciès / perméabilité à chaque maille. Tout d'abord, la valeur y_1 correspondant à une densité de 0,3 de la fonction de répartition est déterminée. Ensuite, chaque maille dans laquelle le champ Gaussien a une valeur supérieure à y_1 est attribuée au faciès 2 (avec une perméabilité de 1000 mD). Les autres sont attribuées au faciès 1 (avec une perméabilité de 100 mD). Le champ de perméabilité de référence est présenté sur la Figure 15 (b).

Cinq puits sont présents dans le réservoir (4 producteurs et 1 injecteur) avec la configuration de la Figure 2. Un débit cible en huile de 20 m^3/jour est imposé aux producteurs, pour un débit d'injection d'eau de 80 m^3/jour. Les données de production (pression de fond, débit d'huile, percée d'eau) sont obtenues en effectuant une simulation d'écoulement sur le modèle de référence pendant 304 jours (10 mois).

3.3.3 Procédure d'assimilation

Les données de production sont assimilées à la fin de chaque mois pendant 10 mois. La distribution de l'erreur sur les mesures a un écart-type égal à 5% de leurs valeurs. Un ensemble de 50 réalisations Gaussiennes centrées réduites est considéré. Le seuil de troncature est appliqué à chacune d'elles afin de générer les distributions en faciès. Le vecteur d'état pour chaque membre d'ensemble j est donc défini par :

- La réalisation de la fonction aléatoire Gaussienne dans chaque maille (variable statique) : $Y_{j,i}$, $i = 1,...,2500$.

- La pression dans chaque maille (variable dynamique) : $P_{j,i}, i = 1,...,2500.$
- La saturation en eau dans chaque maille (variable dynamique) : $Sw_{j,i}, i = 1,...,2500.$
- La pression de fond aux puits : $BHP_{j,k}, k = 1,...,5.$
- Le débit d'huile aux producteurs : $SOR_{j,k}, k = 1,...,4.$
- La percée d'eau aux producteurs : $WCT_{j,k}, k = 1,...,4.$

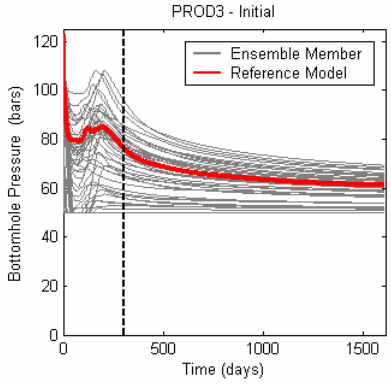
Donc, le vecteur d'état devient :

$$\psi_j = [Y_{j,1}, Y_{j,2}, ..., Y_{j,2500}, P_{j,1}, P_{j,2}, ..., P_{j,2500}, Sw_{j,1}, Sw_{j,2}, ..., Sw_{j,2500}, \\ BHP_{j,1}, ..., BHP_{j,5}, SOR_{j,1}, ..., SOR_{j,4}, WCT_{j,1}, ..., WCT_{j,4}] \quad (15.)$$

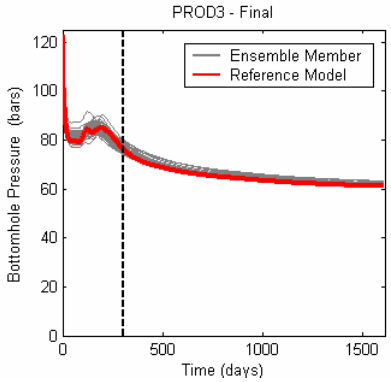
La période de calage est prolongée par une période de prévision de 305 jours.

3.3.4 Résultats du calage d'historique

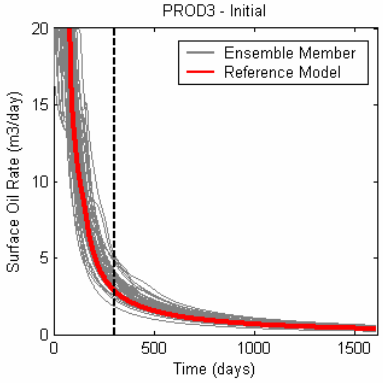
Les données de production simulées avec les ensembles initiaux et finaux sont comparées pour le puits PROD3 sur la Figure 16. Les résultats pour les autres puits sont similaires. On peut observer que les modèles corrigés permettent de simuler des réponses aux puits plus proches des valeurs de référence que les modèles initiaux. L'assimilation entraîne donc une forte diminution de la variation des données simulées autour des valeurs de référence.



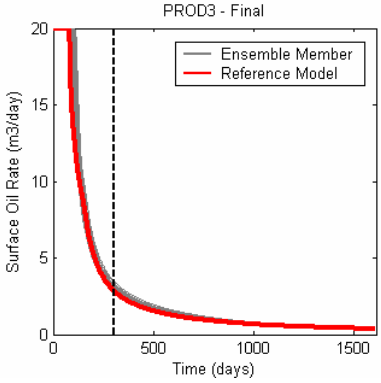
(a) Pression de fond – Ensemble initial



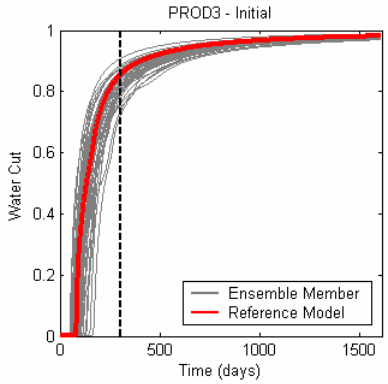
(d) Pression de fond – Ensemble final



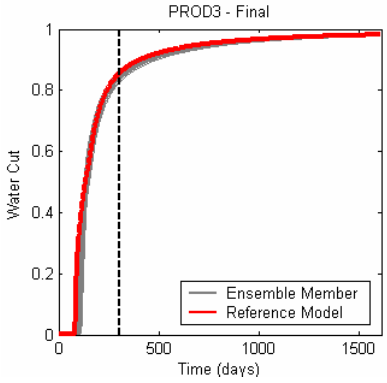
(b) Débit d'huile – Ensemble initial



(e) Débit d'huile – Ensemble final



(c) Percée d'eau – Ensemble initial

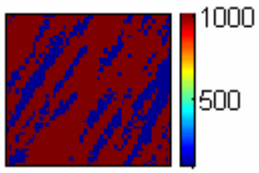


(f) Percée d'eau – Ensemble final

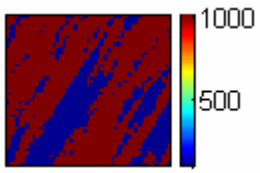
Figure 16 Données de production simulées au puits PROD3 avec les 50 réalisations en faciès de l'ensemble initial (première colonne) et final (deuxième colonne). La ligne verticale noire à 304 jours indique la limite entre la période de calage (0-304 jours) et la période de prévision (305-1612 jours). Les courbes grises correspondent aux membres de l'ensemble. Les valeurs de référence sont tracées en rouge.

L'assimilation de donnée a permis de diminuer le RMS de la pression de fond, du débit d'huile et de la percée d'eau de 14, 61 et 51% respectivement.

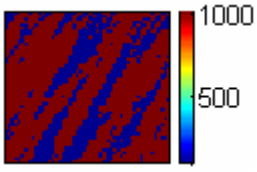
Sur la Figure 17 , on peut voir les distributions de perméabilité initiales et finales pour quatre membres de l'ensemble. Les modèles initiaux ont tous été générés avec la même procédure et apparaissent très différents. Après assimilation, ils présentent des caractéristiques très proches, dont la plupart se retrouvent dans le modèle de référence.



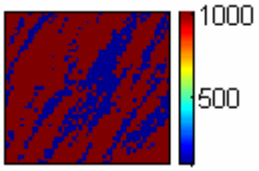
(a) Membre 1 – Initial



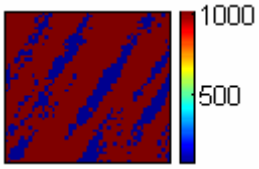
(b) Membre 2 - Initial



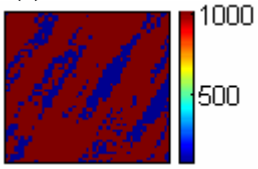
(c) Membre 3 – Initial



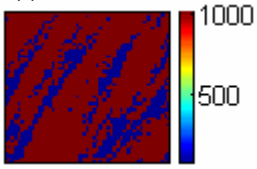
(d) Membre 4 – Initial



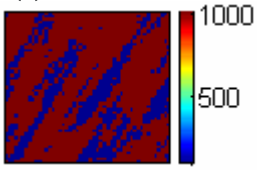
(e) Membre 1 – Final



(f) Membre 2 - Final



(g) Membre 3 – Final



(h) Membre 4 – Final

Figure 17 Valeurs initiales (première ligne) et finales (deuxième ligne) de 4 membres de l'ensemble. Le bleu correspond au faciès 1 (avec une perméabilité de 100 mD) et le rouge au faciès 2 (avec une perméabilité de 1000 mD).

Les distributions mises à jour dans le vecteur d'état au cours des assimilations restent Gaussiennes, mais leur moyenne et variance subissent des changements. Le seuil de troncature utilisé ne permet donc pas de préserver les proportions de facies initiales. Les proportions obtenues pour le faciès 1 au cours des assimilations sont les suivantes :

Tableau 2 Proportion du facies 1 au cours des assimilations

Pas de temps	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Proportion	0.2956	0.2851	0.2802	0.2711	0.2600	0.2770	0.2578	0.2622	0.2544	0.2560	0.2470

4 Méthode EnKF et paramétrisation

Les modèles numériques de réservoir sont constitués de milliards de mailles auxquelles sont associées des propriétés pétrophysiques telles que la porosité et la perméabilité. Une grille plus grossière est généralement utilisée pour les simulations d'écoulement, mais le nombre d'inconnues reste très élevé. Des méthodes de paramétrisation ont donc été développées, qui permettent de perturber le modèle géologique avec un nombre réduit de paramètres tout en conservant ses propriétés spatiales. Parmi elles figurent la méthode des points pilotes (de Marsily, 1984) et la méthode de déformation graduelle (Hu, 2000).

L'objectif de cette partie est de combiner ces deux techniques de paramétrisation avec la méthode EnKF. On cherche ainsi à conserver la variabilité spatiale initiale des modèles au cours des assimilations. Ceci peut améliorer la performance du filtre, comme par exemple éviter le problème de surestimation des paramètres. Dans les algorithmes développés, la méthode EnKF est appliquée à un petit nombre de paramètres statiques permettant de mettre à jour tout le modèle géologique (porosité, perméabilité). Toutefois, ces paramétrisations renforcent la non linéarité des dépendances entre variables d'état et données de production, ce qui peut influer sur la performance de la méthode.

4.1 Algorithme combinant les méthodes EnKF et points pilotes

La méthode des points pilotes permet de perturber la réalisation d'une propriété donnée tout en conservant sa variabilité spatiale. L'idée est de modifier la valeur de cette propriété en certains points, appelés « points pilotes », et de propager ces perturbations à toute la réalisation par krigeage. Si les points pilotes sont suffisamment espacés, la variabilité spatiale définie par le variogramme est préservée. Lors d'un processus de calage d'historique, les valeurs de la propriété aux points pilotes sont considérées comme des paramètres d'inversion et ajustés automatiquement dans la boucle d'optimisation.

4.1.1 Formulation mathématique de la méthode des points pilotes

Considérons une réalisation non conditionnée y d'une propriété donnée et n points pilots. La réalisation y peut être conditionnée à de nouvelles valeurs aux points pilotes par krigeage (Chilès et Delfiner, 1999):

$$y_c(\mathrm{u}) = y_{dK}(\mathrm{u}) + [y(\mathrm{u}) - y_K(\mathrm{u})] \qquad 16.$$

Le vecteur u indique la position, y_{dK} est l'estimateur de krigeage sur les données statiques et les nouvelles valeurs aux points pilotes, et y_K l'estimateur par krigeage sur les valeurs de y aux points pilotes. y_c est une réalisation de la propriété considérée satisfaisant les nouvelles valeurs aux points pilotes.

4.1.2 Algorithme combinant les méthodes EnKF et points pilotes

Le principe de cet algorithme est de ne considérer comme variables d'état statiques que les valeurs des propriétés pétrophysiques aux points pilotes. La méthode complète est résumée sur la Figure 18.

Considérons un ensemble initial, appelé « ensemble de base», contenant N réalisations de chaque propriété statique incertaine générées selon le modèle a priori, et un ensemble de points pilotes. A chaque pas d'assimilation, la méthode se décompose en deux étapes :

1. L'ébauche
    - Chaque réalisation de l'ensemble de base est conditionnée aux valeurs aux points pilotes obtenues à l'assimilation précédente. Les nouvelles réalisations obtenues forment l'ensemble EnKF.
    - Une simulation numérique d'écoulement est effectuée sur chacune des réalisations de l'ensemble EnKF sur la période considérée. Elle permet d'alimenter le vecteur d'état qui contient les valeurs aux points pilotes ainsi que les variables dynamiques et réponses prédites aux puits.
2. La correction
    - L'étape de correction de la méthode EnKF est appliquée au vecteur d'état
    - Les valeurs corrigées des propriétés statiques aux points pilotes et des variables dynamiques forment l'état initial du pas de temps suivant.

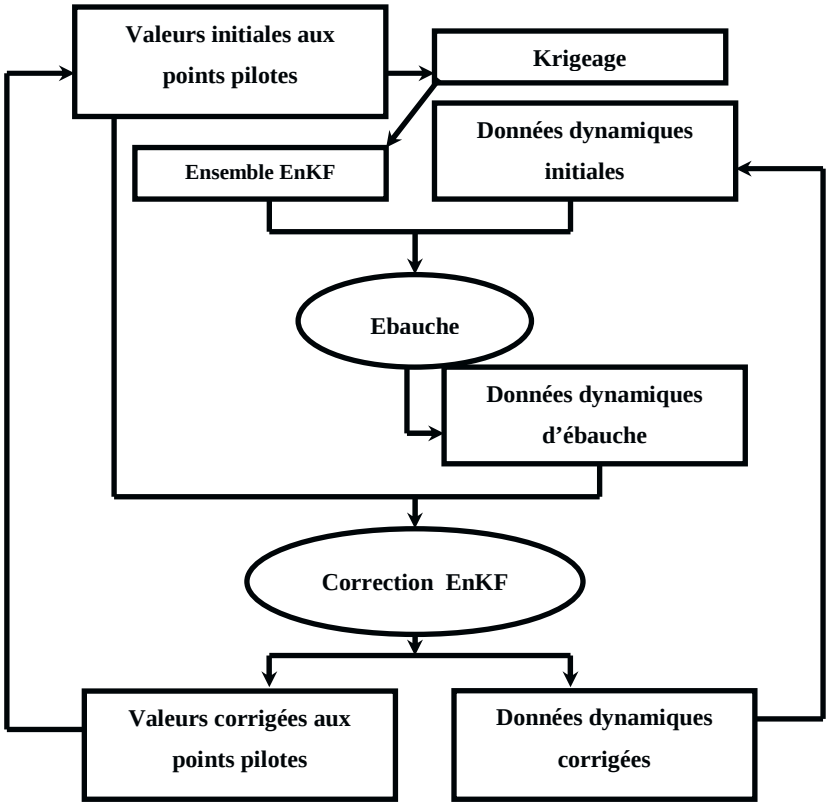


Figure 18 Algorithme combinant les méthodes EnKF et points pilotes

Notons que les valeurs initiales aux points pilotes (première assimilation) sont extraites de l'ensemble de base. Les variables dynamiques correspondent à l'état d'équilibre du réservoir avant production.

4.1.3 Application au calage d'historique

On considère ici le cas 2D simple utilisé dans la section 3.1. 50 réalisations de la perméabilité sont générées pour former l'ensemble de base. Les différentes configurations de placement des points pilotes utilisées sont détaillées dans la section 4.1.4. Le vecteur d'état pour chaque membre de l'ensemble est le suivant :

- Le logarithme de la perméabilité aux points pilotes (variable statique) : $\ln(k_{p,i})$, $i = 1,...,$ Nombre de points pilotes
- La pression dans chaque maille (variable dynamique) : $P_{j,i}, i = 1,...,2500.$
- La saturation en eau dans chaque maille (variable dynamique) : $Sw_{j,i}, i = 1,...,2500.$
- La pression de fond aux puits : $BHP_{j,k}, k = 1,...,5.$
- Le débit de production d'huile aux producteurs : $SOR_{j,k}, k = 1,...,4.$

- La percée d'eau aux producteurs : $WCT_{j,k}, k = 1,...,4.$

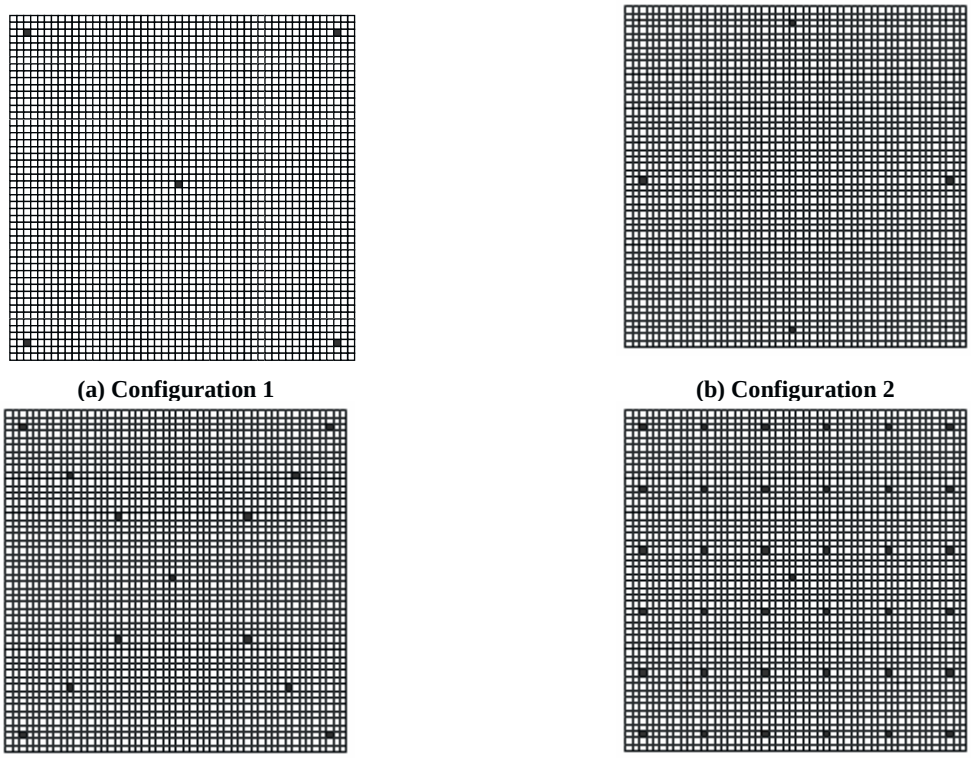
Donc, le vecteur d'état devient :

$$\psi_m = [\ln(k_{p1)},...,\ln(k_{p5}), P_1,...,P_{2500}, Sw_1,...,Sw_{2500}, \quad 17.$$
$$BHP_1,...,BHP_5, SOR_1,...,SOR_4, WCT_1,...,WCT_4]$$

Les périodes de calage d'historique et de prévision sont inchangées.

4.1.4 Positionnement des points pilotes

Quatre scénarios de positionnement des points pilotes sont considérés afin d'examiner la performance de l'algorithme proposé (Figure 19).



(a) Configuration 1 **(b) Configuration 2**

(c) Configuration 3 **(d) Configuration 4**

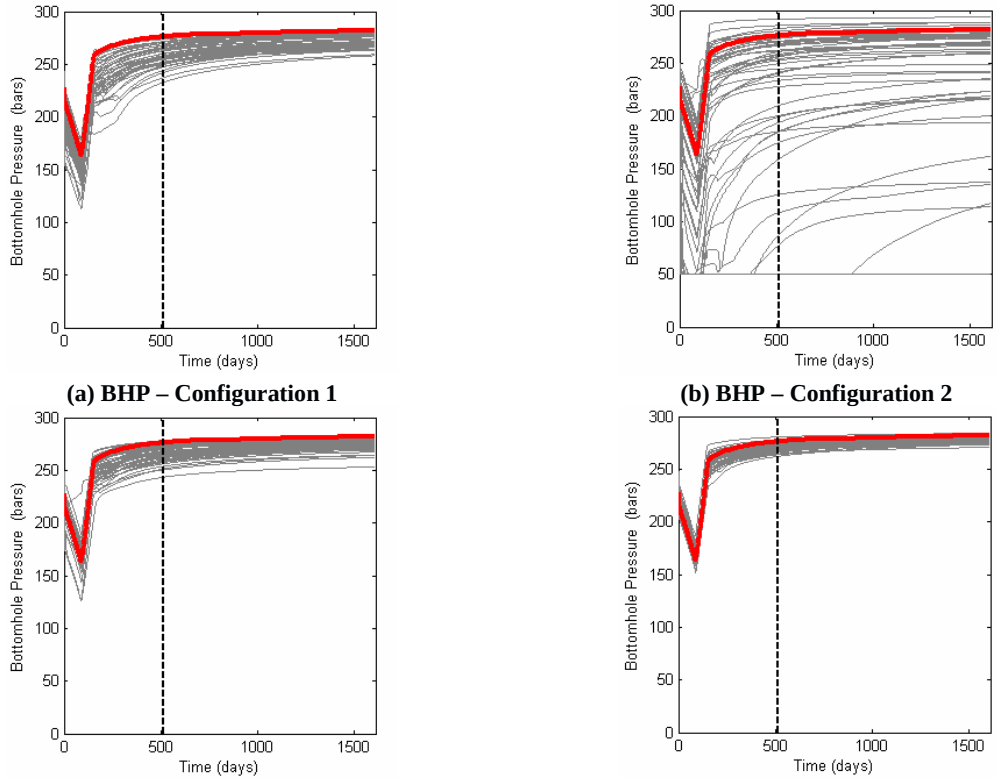
Figure 19 Configurations des points pilotes, représentés par les points noirs.

Dans la première configuration, les points pilotes sont placés aux puits, et sont donc séparés de plus d'une longueur de corrélation du champ de perméabilité. Dans la deuxième configuration, les points pilotes sont placés près des bords entre les puits producteurs, à équidistance de chacun d'eux. Pour obtenir les configurations 3 et 4, des points ont été ajoutées à la première configuration. Dans le cas 3, 13 points sont considérés, situés aux puits ou placés à intervalles réguliers sur les lignes reliant l'injecteur et les producteurs. On cherche ici à capturer les hétérogénéités du milieu dans les directions principales d'écoulement.

Toutefois, la distance entre deux points pilotes peut être inférieure à la longueur de corrélation. La configuration 4 comprend 37 points pilotes : 5 placés aux puits et 32 selon un maillage régulier couvrant tout le réservoir. Cette configuration ne permet pas d'assurer une distance entre deux points supérieure à la longueur de corrélation, mais a été testée afin de voir si l'algorithme était alors capable de régénérer les hétérogénéités à l'échelle fine.

4.1.5 Résultats du calage d'historique

Les résultats de l'assimilation des données de production pour les 4 configurations sont comparés sur la Figure 20 pour la pression de fond (BHP) au puits PROD3. Les valeurs obtenues avec l'ensemble initial sont données sur la Figure 3.



(a) BHP – Configuration 1

(b) BHP – Configuration 2

(c) BHP – Configuration 3

(d) BHP – Configuration 4

Figure 20 Pression de fond au puits PROD3 pour les ensembles finaux : (a) Configuration 1, (b) Configuration 2, (c) Configuration 3 et (d) Configuration 4.

Nous pouvons observer qu'en ajustant les valeurs aux points pilotes des configurations 1, 3 et 4, le calage est amélioré et la dispersion des valeurs simulées autour des données diminue. La configuration 2 ne permet pas de reproduire les valeurs de référence : les points pilotes doivent être placés là où l'écoulement est significatif.

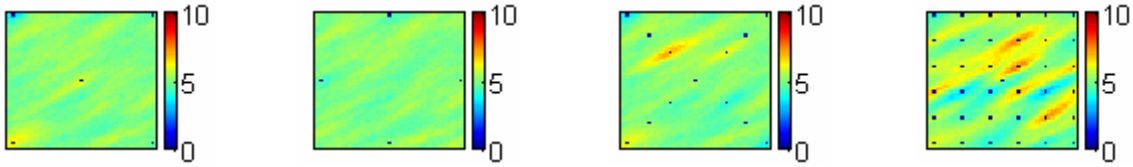
Les ratios de l'erreur RMS des données de production (équation 11) entre l'ensemble final et l'ensemble initial sont donnés dans le Tableau 3. Comme on pouvait s'y attendre, les résultats

les plus mauvais sont ceux de la configuration 2. Le calage s'améliore en ajoutant des points répartis progressivement dans tout le réservoir.

Tableau 3 $RMS_{prod,1}$ pour les quatre configurations de points pilotes

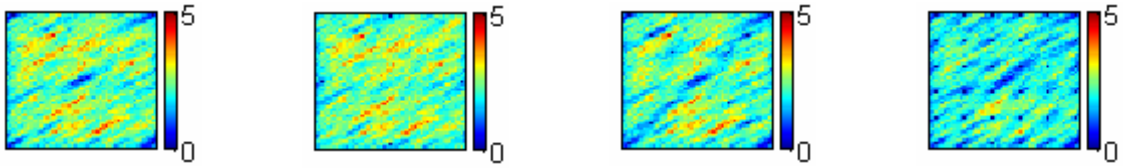
RMS	BHP	OPR	WCT
Configuration 1	0.89	0.88	0.87
Configuration 2	0.99	0.98	0.98
Configuration 3	0.62	0.88	0.89
Configuration 4	0.64	0.82	0.83

La moyenne et la variance de l'ensemble des réalisations du logarithme de la perméabilité sont présentées sur les Figure 21 et Figure 22. En considérant les points pilotes loin des puits comme dans la configuration 2, l'algorithme ne peut pas reproduire les caractéristiques du modèle de référence. Ceci montre que les points pilotes doivent être positionnés là où l'écoulement des fluides est significatif.



(a) Config. 1 – Final **(b) Config. 2 – Fina1** **(c) Config. 3 - Final** **(d) Config. 4 – Final**

Figure 21 Moyenne de l'ensemble des modèles ln(kh) pour les quatre configurations.



(a) Config. 1 – Final **(b) Config. 2 – Fina1** **(c) Config. 3 - Final** **(d) Config. 4 – Final**

Figure 22 Variance de l'ensemble des modèles ln(kh) pour les quatre configurations.

Pour les configurations 1, 2 et 4, les modèles finaux reproduisent quelques caractéristiques du modèle de référence (à grande échelle), particulièrement autour des points pilotes (Figure 21). De plus, la variance des modèles finaux a diminué, notamment autour de ces points (Figure 22). Plus d'hétérogénéités sont visibles sur la moyenne finale de la configuration 4 car on y considère plus de points pilotes placés de manière régulière dans tout le réservoir. La diminution de la variance y est aussi plus importante.

4.2 Méthode permettant de combiner EnKF et déformation graduelle

La méthode de la déformation graduelle a été introduite par Hu (2000). Elle repose sur la combinaison de fonctions aléatoires Gaussiennes ayant les mêmes propriétés statistiques avec

une pondération telle que ces propriétés soient également satisfaites par la fonction aléatoire Gaussienne en résultant. Cette technique permet de perturber toutes les réalisations d'une fonction aléatoire Gaussienne avec peu de paramètres, tout en générant de nouvelles réalisations de cette fonction.

4.2.1 Formulation mathématique de la méthode de déformation graduelle

Considérons deux réalisations y_1 et y_2 de deux fonctions aléatoires Gaussiennes indépendantes Y_1 et Y_2 de même moyenne m et covariance C. Alors, pour tout paramètre t,

$$y - m = (y_1 - m)\cos(\pi t) + (y_2 - m)\sin(\pi t) \qquad 18.$$

définit une réalisation d'une fonction aléatoire Gaussienne Y de moyenne m et de covariance C. t est le paramètre de déformation graduelle. Cette formule est périodique en t sur $-1 \le t \le +1$. Si $t = 0$, y est égal à y_1 ; si $t = 1/2$, y est égal à y_2. Des perturbations continues de t entraînent des changements continus de y.

La déformation graduelle peut être utilisée pour combiner plus de deux réalisations. Considérons M fonctions aléatoires Gaussiennes $\{Y_1, Y_2, ..., Y_M\}$ de moyenne m et covariance C, et un ensemble de réalisations $\{y_1, y_2, ..., y_M\}$ de ces fonctions aléatoires. Alors, pour tout ensemble de M-1 paramètres de déformation graduelle $\{t_1, t_2, ..., t_{M-1}\}$,

$$y - m = \prod_{i=1}^{M-1}(y_1 - m)\cos(t_i) + \sum_{i=1}^{M-1}\sin(t_i)\prod_{j=i+1}^{M-1}\cos(t_j)(y_{i+1} - m) \qquad 19.$$

définit une réalisation y d'une fonction aléatoire Gaussienne de moyenne m et covariance C.

Si la déformation graduelle est appliquée à des bruits blancs Gaussiens, il est possible de réaliser des déformations locales. Dans ce cas, des valeurs différentes des paramètres sont appliquées à des régions séparées du réservoir.

4.2.2 Algorithme combinant la méthode EnKF et la déformation graduelle

L'idée principale de cet algorithme est de considérer comme variable statique dans le vecteur d'état des paramètres de déformation graduelle utilisés pour combiner un ensemble donné de réalisations des propriétés pétrophysiques (porosité, perméabilité), appelé ensemble de base. Chaque ensemble de paramètres définit un modèle réservoir sur lequel est réalisée une simulation d'écoulement. Les distributions des propriétés pétrophysiques mises à jour par le

filtre appartiennent donc à l'espace défini par l'ensemble de base, et respectent ainsi l'information a priori. La déformation graduelle peut être appliquée sur les réalisations de chaque propriété ou sur les bruits blancs Gaussiens utilisés pour générer ces réalisations, par exemple par la méthode FFT-MA.

L'algorithme est schématisé sur la Figure 23. Considérons tout d'abord le cas où la déformation graduelle est appliquée sur les réalisations des propriétés pétrophysiques.

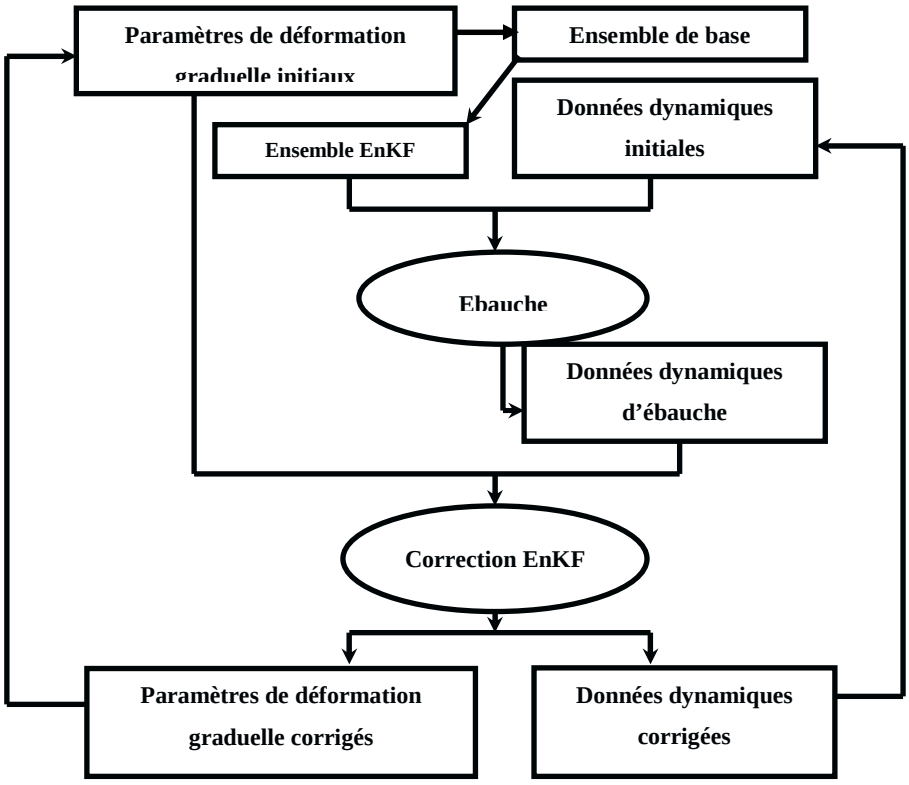


Figure 23 Workflow de l'algorithme combinant la méthode EnKF et la déformation graduelle.

Supposons connus, pour chaque propriété pétrophysique, un ensemble de base contenant M réalisations indépendantes $b_i, i = 1,...,M$ de cette propriété ainsi que N ensembles de $M-1$ paramètres de déformation graduelle $\{t_{j,1}, t_{j,2},...,t_{j,M-1}\}$, $j = 1,...,N$, générés aléatoirement selon une distribution uniforme. Ces paramètres seront considérés comme des variables statiques dans le vecteur d'état après transformation, les variables d'état devant être Gaussiennes. A chaque pas de temps d'assimilation, la méthode se décompose en deux étapes :

1. L'ébauche
    - Les paramètres de déformation graduelle sont appliqués à l'ensemble de base, générant ainsi N réalisations $e_j, j = 1,...,N$ pour chaque propriété pétrophysique :

$$e_j = \prod_{i=1}^{M-1}(b_1)\cos(t_{j,i}) + \sum_{i=1}^{M-1}\sin(t_{j,i})\prod_{k=i+1}^{M-1}\cos(t_{j,k})(b_{i+1}) \quad 20.$$

b_j représente les réalisations centrées réduites de l'ensemble de fond. La formule (20) nous donne donc des réalisations e_j centrées réduites qui doivent ensuite être transformées pour vérifier la moyenne et variance de la propriété. L'ensemble obtenu est appelé ensemble EnKF.

- Une simulation numérique d'écoulement est effectuée sur chaque modèle de l'ensemble EnKF jusqu'au pas de temps d'assimilation suivant. On peut ensuite construire le vecteur d'état de l'ébauche contenant, pour chaque membre de l'ensemble EnKF, les paramètres de déformation graduelle $t_{j,i}, i = 1,...,M-1$ associés, les variables dynamiques prédites à la fin du pas de temps (pression, saturation) et les données de production simulées.

2. La correction
   - L'étape de correction de la méthode EnKF est appliquée au vecteur d'état.
   - Les valeurs corrigées des paramètres de déformation graduelle et des variables dynamiques forment l'état initial pour le pas de temps suivant.

Si la déformation graduelle est appliquée aux bruits blancs Gaussiens, l'ensemble de base est constitué de M bruits blancs Gaussiens $b_i, i = 1,...,M$ pour chaque propriété pétrophysique considérée. Ils sont combinés en appliquant la formule (19) avec les ensembles de paramètres de déformation graduelle $\{t_{j,1}, t_{j,2},...,t_{j,M-1}\}$, $j = 1,...,N$. Les N nouveaux bruits blancs ainsi créés sont ensuite utilisés pour générer les réalisations $e_j, j = 1,...,N$ de la propriété, par exemple en appliquant l'algorithme FFT-MA.

Dans certains cas, la déformation graduelle globale peut manquer d'efficacité pour le calage d'historique. Par exemple, l'amélioration du calage à un puits peut entraîner sa détérioration à un autre puits. Dans ce cas, la déformation graduelle locale peut apporter plus de flexibilité et améliorer le calage. Cela nécessite d'appliquer la déformation sur les bruits blancs.
Supposons que la grille réservoir soit divisée en Z zones contenant $Num_k, k = 1,...,Z$ mailles telles que :

$$\sum_{k=1}^{Z} Num_k = \text{ Nombre total des mailles} \quad 21.$$

Considérons un ensemble EnKF de taille N et un ensemble de base de taille M. N ensembles de $M-1$ paramètres de déformation graduelle $\{t_{j,1}^k, t_{j,2}^k, ..., t_{j,M-1}^k\}$, $j = 1,..., N$ sont alors générés pour chaque zone $k = 1,..., Z$ et utilisés pour combiner les bruits blancs Gaussiens de l'ensemble de base $b_i, i = 1,..., M$ selon la formule :

$$\tilde{b}_j(x) = \prod_{i=1}^{M-1} b_1(x)\cos(t_{j,i}^{k(x)}) + \sum_{i=1}^{M-1} \sin(t_{j,i}^{k(x)}) \prod_{l=i+1}^{M-1} b_{i+1}(x)\cos(t_{j,l}^{k(x)}) \qquad 22.$$

où on a noté x l'indice des mailles de la grille et $k(x) \in \{1..Z\}$ l'indice de la zone à laquelle la maille x appartient. Les bruits blancs obtenus $\tilde{b}_j$ servent ensuite à générer les réalisations e_j, $j = 1,..., N$ de la propriété pétrophysique.

4.2.3 Application de la méthode pour le calage d'historique

Le cas d'application considéré ici est le cas synthétique 2D décrit dans la section 3.1. Cette expérience est considérée en tant que le cas de base de cette section. La méthode FFT-MA est utilisée pour générer les réalisations du logarithme de la perméabilité selon les caractéristiques du modèle de référence.

Le vecteur d'état pour chaque membre de l'ensemble, j, est donc :

- L'ensemble des M-1 paramètres de déformation graduelle (variable statique) : $t_{j,i}$, $i = 1,..., \text{M-1}$.
- La pression dans chaque maille (variable dynamique) : $P_{j,i}, i = 1,...,2500$.
- La saturation en eau dans chaque maille (variable dynamique) : $Sw_{j,i}, i = 1,...,2500$.
- La pression de fond aux puits : $BHP_{j,k}, k = 1,...,5$.
- Le débit d'huile aux producteurs : $SOR_{j,k}, k = 1,...,4$.
- La percée d'eau aux producteurs : $WCT_{j,k}, k = 1,...,4$.

Donc, le vecteur d'état devient :

$$\psi_m = [t_{m,1},...,t_{m,M-1}, P_{m,1},...,P_{m,2500}, Sw_{m,1},...,Sw_{m,2500}, \\ BHP_{m,1},...,BHP_{m,5}, SOR_{m,1},...,SOR_{m,4}, WCT_{m,1},...,WCT_{m,4}] \qquad 23.$$

Les périodes de calage et de prédiction sont inchangées.

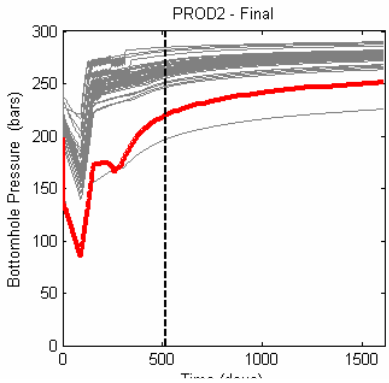
4.2.4 Résultats du calage d'historique

Trois séries d'expériences sont présentées dans cette section : la première considère la déformation graduelle appliquée aux bruits blancs, la seconde consiste à appliquer des déformations graduelles locales, et la dernière traite de la déformation graduelle appliquée aux réalisations de ln(kh).

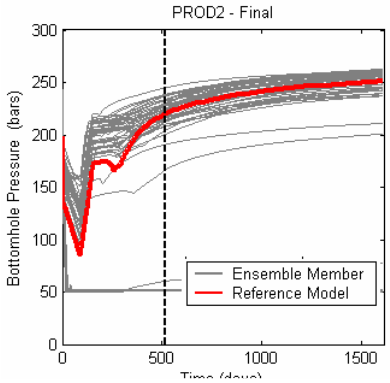
4.2.4.1 Déformation graduelle appliquée aux bruits blancs Gaussiens

En augmentant la taille de l'ensemble, on élargit l'espace de recherche pour les modèles statiques mis à jour pendant l'assimilation. De plus, la performance des méthodes de paramétrisation augmente généralement avec le nombre de paramètres considérés.

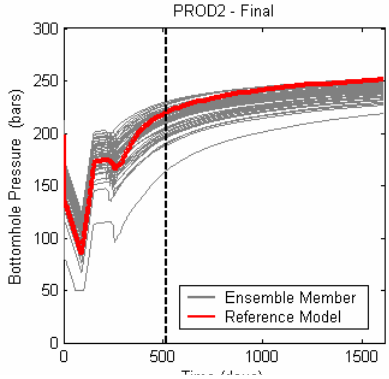
Nous considérons ici un ensemble EnKF de taille N=100, des ensembles de base de taille $M = 100$, 1000, et 2500, ces derniers étant composés de bruits blancs Gaussiens. Les autres paramètres de la méthode sont les mêmes que dans la section 3.1. Les pressions de fond simulées au puits PROD2 avec les modèles finaux sont présentées sur la Figure 24. On peut voir que les valeurs simulées sont d'autant plus proches des valeurs de référence que la taille de l'ensemble de fond est élevée. Cependant, la performance de la méthode EnKF traditionnelle est supérieure à celle de ce nouvel algorithme.



(a) BHP – Ensemble Final
Taille de l'ensemble de base : 100



(b) BHP – Ensemble Final
Taille de l'ensemble de base : 1000



(c) BHP – Ensemble Final
Taille de l'ensemble de base : 2500

Figure 24 Pression de fond au puits PROD2 pour les ensembles finaux, en considérant des ensembles de base de taille (a) 100, (b) 1000 et (c) 2500. La ligne pointillée noire à 516 jours indique la limite entre la période de calage (0-516 jours) et la période de prévision (517-1612 jours). Les courbes grises correspondent aux membres de l'ensemble et la courbe rouge aux valeurs de référence.

L'erreur $RMS_{prod,1}$ calculée pour les ensembles finaux des trois expériences ci-dessus est donnée dans le Tableau 4, normalisée par la valeur obtenue avec un ensemble de base de taille 100. On peut donc voir que cette erreur diminue fortement lorsque l'on augmente la taille de l'ensemble de base : l'algorithme apparaît plus performant avec un espace de recherche plus large.

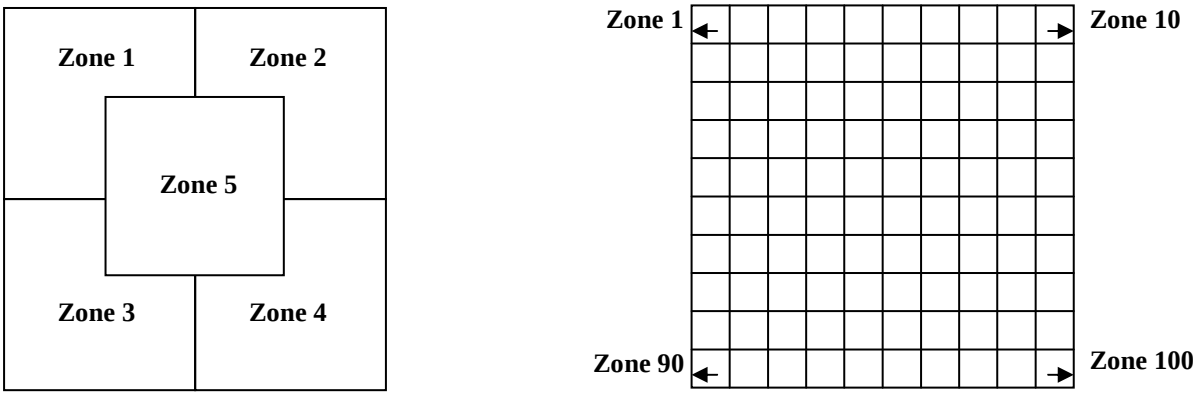
Tableau 4 Effet de la taille de l'ensemble de base, la déformation graduelle étant appliquée aux bruits blancs Gaussiens. Les valeurs du $RMS_{prod,1}$ sont normalisées par celles obtenues avec un ensemble de base de taille 100.

Taille de l'ensemble de base	BHP	SOR	WCT
100	1.0	1.0	1.0
1000	0.7390	0.7764	0.7494
2500	0.3593	0.4124	0.4134

Enfin, la ressemblance des modèles finaux de perméabilité avec le modèle de référence augmente avec la taille de l'ensemble de base, même s'il n'est pas possible de reproduire toutes ses caractéristiques.

4.2.4.2 Déformation graduelle locale

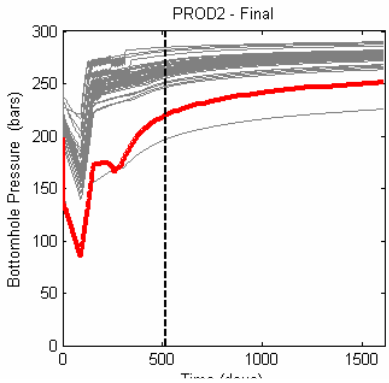
On applique ici la déformation graduelle locale aux bruits blancs Gaussiens en considérant deux découpages de la grille réservoir, et des ensembles de taille N=50 et M=100. Dans la première configuration (Figure 25 (a)), le réservoir est divisé en cinq zones non régulières autour de chaque puits. Les zones 1 à 4 contiennent 481 mailles et la zone 5, 576 mailles. Dans la deuxième configuration (Figure 25(b)), la grille est divisée en régions identiques de 25 mailles chacune.



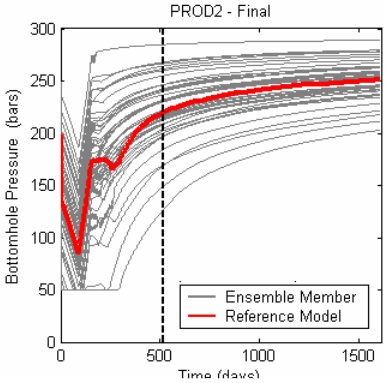
(a) Configuration à 5 zones (b) Configuration à 100 zones

Figure 25 Configuration des zones pour la déformation graduelle locale (a) Configuration à 5 zones, (b) Configuration à 100 zones.

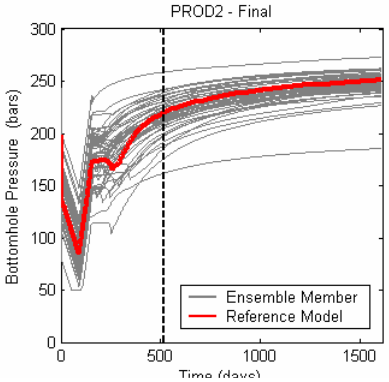
La Figure 26 présente les pressions de fond simulées au puits PROD2 avec les modèles finaux obtenus avec la déformation graduelle globale (1 zone) et les deux configurations locales. En augmentant le nombre de zones, ces pressions sont mieux calées. La configuration à 5 zones donne ici les meilleurs résultats pour le débit d'huile et la percée d'eau. Ceci peut être lié au fait que l'ensemble EnKF initial n'est pas le même pour les trois expériences.



(a) BHP – Ensemble Final - 1 zone



(b) BHP – Ensemble Final - 5 zones



(c) BHP – Final Ensemble – 100 zones

Figure 26 Pression de fond au puits PROD2 simulée avec les modèles finaux obtenus avec la déformation graduelle locale (a) 1 zone, (b) 5 zones et (c) 100 zones. La ligne pointillée noire à 516 jours indique la limite entre la période de calage (0-516 jours) et la période de prévision (517-1612 jours). Les courbes grises correspondent aux membres de l'ensemble et la courbe rouge aux valeurs de référence.

L'erreur $RMS_{prod,1}$ des données de production (équation 11) obtenue pour les ensembles finaux est donnée dans le Tableau 5. Quand le nombre de zones augmente, le modèle de référence est mieux calé.

Tableau 5 Effet de la déformation graduelle locale avec trois configurations de zones. Les valeurs du $RMS_{prod,1}$ sont normalisées sur celles obtenues avec une seule zone (déformation globale).

Nombre de zones	BHP	SOR	WCT
1	1.0	1.0	1.0
5	0.8049	0.6781	0.6770
100	0.1877	0.5770	0.5762

4.2.4.3 Déformation graduelle appliquée aux réalisations pétrophysiques

On applique ici l'algorithme proposé aux réalisations de ln(kh), en considérant un ensemble EnKF de taille 50 et des ensembles de base de taille 100, 1000 et 2500. Comme précédemment, augmenter la taille de l'ensemble de base permet de mieux caler les données de production. Cependant, les performances de la méthode EnKF traditionnelle restent supérieures à celles de l'algorithme de combinaison. Cela peut être expliqué par le fait que la paramétrisation rend le problème encore plus non linéaire.

Le Tableau 6 présente l'erreur $RMS_{prod,1}$ pour les données de production obtenue en augmentant la taille de l'ensemble de base. Les valeurs sont normalisées sur celles du cas M=100. Comme précédemment, les données de production apparaissent mieux calées pour des ensembles de base plus larges.

Tableau 6 Effet de la taille de l'ensemble de base, la déformation graduelle étant appliquée aux réalisations. Les valeurs du $RMS_{prod,1}$ sont normalisées sur celles de l'ensemble de taille 100.

Taille de l'ensemble de base	BHP	SOR	WCT
100	1.0	1.0	1.0
1000	0.7452	0.8740	0.8757
2500	0.6211	0.7740	0.7740

5 Conclusions

Dans ce travail, nous nous sommes intéressés à la dernière étape de l'étude d'un réservoir pétrolier : le calage d'historique. Il s'agit de contraindre le modèle de ce réservoir aux données dynamiques collectées sur le terrain, comme les mesures aux puits des pressions, débits ... Plus précisément, ces travaux portent sur l'utilisation de la méthode du filtre de Kalman d'ensemble (EnKF) pour réaliser ce calage. Cette méthode présente l'avantage d'intégrer les

données de production dès qu'elles ont été acquises sans avoir besoin de re-simuler les périodes d'assimilation précédentes. Toutefois, elle repose sur des hypothèses spécifiques dont le non-respect peut entraîner une dégradation de ses performances.
Ce travail s'articule autour de deux objectifs principaux: examiner les performances de la méthode pour le calage d'historique au travers de plusieurs cas d'étude synthétiques ; et présenter de nouvelles méthodes permettant de combiner la méthode EnKF avec des techniques de paramétrisation permettant de préserver la variabilité spatiale des modèles initiaux au cours de l'assimilation des données.

Dans un premier temps, nous avons appliqué la méthode EnKF pour le calage de données de production sur trois cas d'étude synthétiques. Dans le premier, le logarithme du champ de perméabilité horizontale est contraint aux données de pression, débit d'huile et percée d'eau mesurées aux puits. Le second est une variante du cas PUNQ-S3, considérant à la fois la porosité et la perméabilité comme des paramètres incertains. Enfin, le dernier cas traite du problème des modèles discrets en faciès.
Dans chaque cas, l'assimilation des données de production avec la méthode EnKF a permis de réduire la dispersion parmi les modèles de l'ensemble, et la plupart d'entre eux sont capable de régénérer les données de production de modèle de référence pendant les périodes de calage et de prévision. Les distributions des propriétés pétrophysiques après corrections reproduisent certaines caractéristiques des modèles de référence. Enfin, des tests de sensibilité effectués sur les deux premiers cas ont mis en évidence les points suivants :

- La taille de l'ensemble est un paramètre très influent : ici, une taille de 100 semble appropriée. Un ensemble plus grand peut améliorer les résultats mais l'effet reste limité. Un ensemble de taille plus petite que 100, 50 par exemple, peut entraîner une diminution significative de la variance au sein de l'ensemble corrigé, qui apparaît alors inapproprié pour faire des prévisions.
- De petits pas de temps d'assimilation peuvent être nécessaires pour capturer des changements abrupts dans l'écoulement des fluides, par exemple à l'ouverture ou à la fermeture d'un puits.
- Le choix du niveau d'incertitude sur les valeurs mesurées n'a que peu d'effet sur les résultats d'assimilation.
- Enfin, une petite taille d'ensemble peut entraîner une surestimation des paramètres, tels que la porosité pour le cas d'étude PUNQ-S3.

Enfin, le cas en faciès a mis en évidence le fait que le calage des données de production était satisfaisant, mais les proportions de faciès n'étaient pas préservées.

Dans la deuxième partie de ce travail, deux nouveaux algorithmes ont été proposés, qui permettent de combiner la méthode EnKF avec la technique des points pilotes d'une part, et la déformation graduelle d'autre part. Le but est ici de réaliser des assimilations par le filtre de Kalman d'ensemble tout en préservant les propriétés statistiques a priori d'ordre 2 pour les distributions des propriétés pétrophysiques. Ces algorithmes de combinaison reposent sur l'utilisation d'un ensemble de réalisations "de base" perturbé après chaque assimilation selon la valeur corrigée de paramètres associés: les valeurs aux points pilotes ou les paramètres de déformation graduelle. Ces derniers sont donc considérés comme des variables statiques dans le vecteur d'état à la place des distributions des propriétés pétrophysiques dans le réservoir.
La performance de ces algorithmes a été testée sur le premier cas d'étude. Dans le cas des points pilotes, elle peut être améliorée en augmentant le nombre de points pilotes ainsi que leur positionnement. Cependant, la distance entre deux de ces points ne doit pas dépasser une longueur de corrélation.
L'algorithme combinant la méthode EnKF et la déformation graduelle apparaît plus contraignant que la méthode EnKF traditionnelle dans le sens où les modèles corrigés appartiennent nécessairement à l'espace défini par l'ensemble de base de taille finie. De plus, l'utilisation des paramètres de déformation graduelle semble accroître la non-linéarité avec les réponses simulées par rapport à la méthode EnKF traditionnelle. Toutefois, la performance de l'algorithme peut être améliorée en utilisant un ensemble de base plus grand ou en appliquant la déformation localement. La dépendance des résultats avec l'ensemble de base apparaît très importante.

Les perspectives envisagées pour la suite de ce travail sont :

- L'application de la méthode EnKF à un cas d'étude réel ;
- L'application de la méthode EnKF avec des techniques de localisation pour améliorer les résultats sur la variante du cas PUNQ-S3.
- L'application des algorithmes combinant les méthodes EnKF et points pilotes / déformation graduelle à la variante du cas d'étude PUNQ-S3 et au cas en faciès.
- La recherche d'autres algorithmes permettant de combiner EnKF et déformation graduelle.

Table of Contents

List of Tables

List of Figures

Abstract

Due to the extent of petroleum reservoirs and the complexity of the geological environments, direct measurements of petrophysical properties such as porosity and permeability are limited. Thus, the initial reservoir model is constructed with a certain degree of uncertainty. As soon as production starts, well production responses such as bottom hole pressure, surface oil rate and water production rate are obtained. These data are indirectly related to the reservoir properties, such as porosity and permeability. However, their integration in reservoir models decreases the uncertainties in these models. This process is called “history-matching”. There exist several history-matching methods in the petroleum literature.

The main method in this research project is the “Ensemble Kalman Filter (EnKF)” method proposed by Evensen (1994). This method extends the Kalman filter (Kalman, 1960) and makes it possible to perform data assimilation with nonlinear dynamic systems. Usually, in any Kalman filter, the time study is split into several time steps, each of them being associated to data acquisition instants. During each time step, the model is first “forecasted” according to the dynamic model operator. Then, the forecasted model is “updated” to account for the newly collected measurements. This updated model is considered as the initial model of the following time step. The update is linear and calls for the estimation of the error covariance matrices related to model parameters and variables at the target time. When handling a nonlinear dynamic model, these error covariances can be approximated from a finite size ensemble (or group) of possible realizations for model parameters and variables. This is the cornerstone of the EnKF. During each time step, the model realizations are all forecasted through time and their uncertainties are derived from averages over the ensemble. During the update step, each model is updated according to the measurements at that time step. Then, it yields a more accurate model for the following time step.

This method has been recently applied in many fields involving inverse problems. History-matching in reservoir engineering is one of these applications. It consists of identifying a geological model populated by petrophysical properties from production data. Following this method, production data are integrated sequentially in the reservoir model.

This methodology takes advantage of the properties listed hereafter.

- The method is fairly simple and its implementation is straightforward.

- It copes with any fluid flow simulator and does not require to modify the simulator's source code.
- The EnKF is a real-time updating method: as soon as new data are obtained, the reservoir model is updated. There is no need to restart the matching study from scratch.
- This method results in an ensemble of updated models, which can be used to assess uncertainties.

However, some drawbacks can also be pointed out.

- Theoretically, the method requires an infinite ensemble size. In practice, we consider only a limited finite ensemble.
- Theoretically, the method requires Gaussian distributions for model parameters and variables. Again, in practice, nature is rarely Gaussian.
- Theoretically, an approximately linear relation between model parameters/variables and measurements is required. Flow simulation is a nonlinear problem.

This research project is the first EnKF experience at "IFP Energies nouvelles". It was conducted in collaboration with the Geostatistics Group of the MINES-ParisTech center. The purpose is to clearly identify the practical merits and disadvantages of the methodology through various applications. The cases studied in both the classical EnKF framework and some modified versions of it are listed below.

- Traditional EnKF application:
  - A two-dimensional synthetic case: the aim is to improve the estimates of a permeability field by matching well production data. The case study is rather simple and does not include any strong changes in fluid flow conditions. The effect of several controlling parameters such as the ensemble size, the assimilation time interval, the uncertainty in measurements, the accuracy of the initial geostatistical properties and the choice of the initial ensemble are studied through sensitivity tests (Chapter 3).
  - A three-dimensional synthetic case study (a variant of the PUNQ-S3), which is usually used as a bench mark in history-matching studies: the purpose is to investigate the shortcomings of the EnKF method whenever the flow behavior is driven by strongly varying flow conditions such as sudden and significant changes in injection and production rates, which enhance non linearity effects (Chapter 4).
  - A simple facies case study with given proportions. Facies models cannot be updated directly by the EnKF due to their non-continuous nature. Therefore, an

intermediate parameterization technique, the truncated Gaussian method is used to model and update facies. The aim is to assess the performance of the EnKF when dealing with non-continuous models (Chapter 5).

- Variants of the EnKF method which call for a link with intermediate parameterization methods. We actually developed such methods in order to make sure that the update step of the EnKF yields reservoir models whose petrophysical properties still respect the prior two-order statistics. The two parameterization techniques considered are the pilot point (Chapter 6) and the gradual deformation (Chapter 7) methods. In both cases, the developed algorithms are applied to the two-dimensional case study of Chapter 3.

Each of the applications and newly developed algorithms are detailed in the corresponding chapters. The performance of the EnKF or the newly developed algorithms in combination with the EnKF are assessed through several experiments. In cases where difficulties are encountered, alternatives are suggested.

1 Introduction

The aim of this chapter is to introduce reservoir modeling and conditioning. The reservoir characterization problem and its different steps are briefly presented. The core of this work, presented as the last step in this reservoir modeling workflow, is “history-matching”. To go through this problem, we refer to the "Ensemble Kalman Filtering (EnKF)".

1.1 Reservoir characterization

The study of a petroleum reservoir is a multidisciplinary work, which calls for various geosciences competences (geology, geophysics, reservoir). The final purpose is to build the most representative model of a reservoir in terms of production forecast, uncertainty quantification and petroleum resources analysis.

The development of oil industry, like many other industries, is motivated by an end goal: the economical interests in addition to supplies and demands. Therefore, industry's history is marked by strong efforts to determine a better description of reservoirs through reservoir models. These models are used to forecast resources and reserves. In addition, they make it possible to identify the best production scenario. Up to now, this motivated the development of simulation workflows, which provide reservoir models consistent with all the data collected on the field: log data, measurements performed on cores extracted from wells, seismic, production data, *etc*. The main steps of this workflow (Figure 1-1) are recapped hereafter:

- Geological Modeling
- Upscaling
- Fluid Flow Simulation
- History-Matching

A general overview is provided about each of these modeling steps in the following sections so that the reader gets introduced to the basics of reservoir modeling.

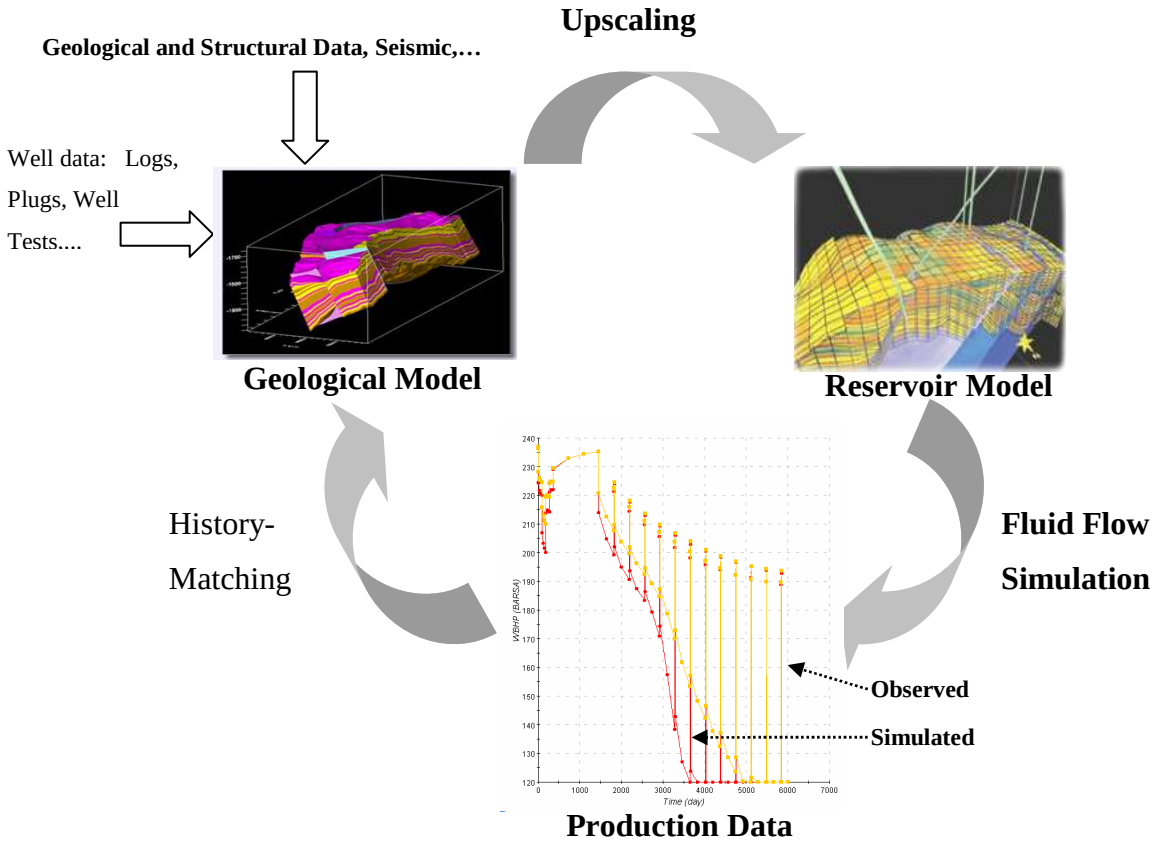


Figure 1-1 Standard workflow in building reservoir models.

1.1.1 Geological model

Initially, an underground geological formation is specified as a potential petroleum bearing area based upon preliminary studies such as geological and seismic surveys. This petroleum reservoir (oil or gas) can extend over hundreds of kilometers by hundreds of kilometers while its thickness is roughly a few hundreds of meters. Moreover, it is burdened deep in the ground and is usually difficult to reach. Therefore, it is required to perform several indirect tests on this area and process the results to access reservoir properties. The geological data about this area are acquired mainly through well logs and cores. During well logging, a measurement probe is descended through the well and measurements are performed that allow for determining lithology, petrophysical properties and fluid types around the well. Meanwhile, samples of reservoir rock, containing reservoir fluid, are extracted and kept under bottom hole conditions. These core samples are then tested in laboratories and provide detailed information about petrophysical properties such as porosity, permeability and relative permeability. Based upon well logs, core samples, existing depositional, sedimentological and

stratigraphical models and analog outcrops, a geologist describes the structure of the petroleum bearing area and provides spatial characteristics describing this area and its petrophysical properties. These characteristics are later used by a geologist to build a numerical "geological model", which is a grid system populated by petrophysical properties. The spatial distributions of these properties are usually described through two-order statistics. A variogram model is defined from the data collected at wells to characterize the spatial variability of these properties. Then, it is used to define a random function. At this stage, the geologist can draw equally probable numerical models, called realizations, using geostatistical simulation algorithms. The resolution, *i.e.* grid volume, chosen for these numerical models depends on the spacing between the finest measured data: that is the spacing between cores or logs. It is usually of the order of 20 cm^3 to 20000 cm^3 (Le Ravalec-Dupin, 2005). This fine scale is inappropriate for any further processing in terms of computation because it would be too CPU-time demanding. Therefore, the fine scale geological model is "upscaled" to a coarser gridded model that we refer to, from now on, as a "reservoir model".

1.1.2 Upscaling

Upscaling consists of aggregating several fine grid blocks to a coarser grid block. The main issue with upscaling is how the petrophysical properties of the coarse grid should be derived from the values of the fine grid and this issue depends on the nature of the property considered. For example, porosity is the ratio of pore volume to total rock volume. This property is additive: the coarse grid block porosity is derived from the arithmetic average of the porosity values of the fine grid blocks included in the coarse grid block. This problem gets more complicated for permeability, which characterizes the ability of a fluid to flow in a porous rock. Referring to equivalence criteria, it is usually required that the flow rate (Rubin and Gomez-Hernandez, 1990) or the dissipated energy (Indelman and Dagan, 1993) should be the same for the group of fine grid blocks and the coarse grid block used instead of the fine ones. Thus, we can estimate an equivalent permeability for the coarse grid block. There exist several approaches to calculate the equivalent permeability, which are split into two groups: the algebraic methods and the numerical methods. A complete discussion can be found in Le Ravalec-Dupin (2005). The upscaled properties populate the coarse reservoir model, which is used to simulate fluid flow.

1.1.3 Reservoir Model and Fluid Flow Simulation

The main question in a reservoir study is how the reservoir will behave when production starts. To handle this issue, we refer to the "reservoir numerical model" which consists of a coarse grid system populated by the properties necessary to track the flow of fluids in the reservoir. These properties range from petrophysical properties such as porosity and permeability to the existing fluids in the reservoir, their fractional phases and contacts, compressibility features, *etc.*

Another classification divides these properties into two groups: the static one and the dynamic one. Static variables do not change with time (as suggested by the term “static”) and originate from geology, well logs, core analysis and seismic surveys. Thus, the petrophysical properties mentioned above are tagged as static variables. We will come back later on to the definition of dynamic variables.

There exist analytical solutions describing the physics of fluid flow in porous media, for a simplified reservoir model saturated by basic (oil and water) fluids (Cossé, 1993). However, a reservoir is mainly heterogeneous, which makes the use of analytical solution unfeasible. As a result, fluid flow is estimated numerically using a "reservoir flow simulator". In other words, a reservoir flow simulator is the engineer's tool to deal with the dynamics of a petroleum reservoir. In general, it solves flow equations for given boundary conditions using a finite difference algorithm. More details are available in Cossé (1993) and Le Ravalec-Dupin (2005). Finally, the fluid flow simulator provides various responses related to fluid displacements in the reservoir and at wells: they are crucial for any reservoir characterization and prediction purposes.

1.1.4 History-Matching

The only type of data available before producing a reservoir is the static data set. These data are linearly related to the reservoir properties considered to populate the reservoir model. Their integration in the reservoir model is thus straightforward. However, due to the extent of the reservoir and its heterogeneous nature, the initial model still requires improvements or refinement to be able to provide reliable production forecasts.

As soon as production starts, the gauges measuring pressures and rates at wells monitor the dynamical behavior of the reservoir. These data are also said dynamic as they vary with time. They are generally subdivided into two main categories. The first one encompasses the

properties attributed to each grid block such as pressure or fluid saturations and the second one the properties related to production or injection wells such as production or injection rate, bottom hole pressure or water cut. In this text, the first subcategory is mainly referred to as dynamic variables and the second one as measurements, observations or production data. Dynamic data are an important source of information for improving the existing reservoir model. However, integration of these data in reservoir model building is complex due to the nonlinear dependence between these data and the modeled properties. The process that incorporates the measurement data into the reservoir simulation workflow is known as "history-matching" in the petroleum literature. In this process, descriptive reservoir parameters, such as petrophysical properties (porosity and permeability), aquifer activity coefficients, fracture locations, fluid compositions, *etc*, are successively adjusted until the updated reservoir model reproduces as well as possible the measured or observed production data.

History-matching has been extensively studied for a long period: the involved challenges are the mathematical complexity, computational cost and the non-uniqueness of results. There exist two approaches for history-matching: manual and assisted or automated. In the first approach, the reservoir descriptive parameters are changed by trial and error until an acceptable match is obtained. This renders the process tedious and may result in unrealistic updated reservoir models, for example with respect to geological constraints. In assisted history-matching, the manual work is decreased and the likelihood of the matched geological models is improved. In other words, the aim is to provide a better estimation of uncertain descriptive reservoir parameters while preserving the initial geological consistency and reproducing the observed production data. In general, the process consists in improving an initial reservoir model by integrating measurement data in a closed optimization loop. At the end of each loop, the departure of the updated reservoir model from the initial geological constraints and the misfit between the simulated and observed production data are evaluated. Then, the loop may be repeated depending on the quality expected for the match. In this case, a new set of reservoir properties is automatically chosen by the optimization algorithm.

The complexity of the optimization problem and the computation cost motivated many works in assisted history-matching. Variational history-matching, which rely on the minimization of an objective function measuring the suitability of the reservoir model considered, involve simultaneously the whole set of production data. When new data are available, the already matched model is refined by restarting the minimization process based upon a new objective function with the previously collected data plus the new ones. The "Ensemble Kalman Filter

(EnKF)" method relies on a different philosophy: data are sequentially integrated in models. This means that flow simulations are not restarted from time 0, but from the last measurement time. This feature makes the EnKF appealing for dealing with systems continuously monitored.

The Ensemble Kalman Filter, initially proposed by Evensen (1994), is a Monte Carlo approach, in which an ensemble of model parameters (petrophysical properties) and states (dynamic variables) is sequentially updated when new data are collected. This makes the method suitable for real-time reservoir history-matching. Its implementation turns out to be simple and independent of the reservoir simulator. In addition, the final multiple matched models are suitable for handling the non-unique nature of history-matching. They can contribute to estimation of uncertainty in the predictions of the model.

1.2 Research objectives

The objective of this work is to investigate the ability of the EnKF method to perform history-matching through various numerical experiments. There have been various applications of this method in different areas of petroleum engineering. In this regard, we aim to investigate the method for reservoir characterization, highlight its merits and weaknesses, introduce the existing improvement techniques and propose new ones.

The main outlines of this work are recapped as follows:

1) Comprehensive study of the state of the art in the "Ensemble Kalman Filtering".
2) Application of the EnKF to several synthetic case studies in order to
   - assess the quality of the match for well production data given uncertain petrophysical properties;
   - investigate the potential of the final updated parameters to forecast production and quantify uncertainty;
   - perform sensitivity studies on the major parameters in the EnKF algorithm such as the choice of the initial ensemble, ensemble size, measurement uncertainty and assimilation time steps.
   - highlight some shortcomings of the method.
3) Application of the EnKF to a synthetic facies model in order to improve facies modeling.

4) Combination of the EnKF with two parameterization methods to preserve the spatial variability of the initial geological model and application to a simple synthetic case study. The considered parameterization methods are:
   - the pilot point method;
   - the gradual deformation parameterization method.

1.3 Outline of chapters

In this work, we investigate the potential of the EnKF for performing assisted history-matching. Chapter 2 presents the theoretical framework for the ensemble Kalman filtering and shows how it can be applied to petroleum reservoir studies. Then, the weaknesses of the traditional EnKF are briefly outlined and the solutions already proposed in the literature are listed. The following chapters are divided into two groups: the first one deals with the application of the traditional EnKF to several case studies and the second one considers two newly developed algorithms combining the EnKF with specific parameterization methods.

The first group comprises Chapters 3 to 5. In Chapter 3, the EnKF is applied to a simple two-dimensional, two phase flow synthetic case study to assess the efficiency of the EnKF for history-matching. In addition, we investigate the prediction capabilities of this method. Last, we study the sensitivity of the EnKF to different parameters. Chapter 4 closely follows Chapter 3 although it involves a more complex case, consisting of a 2 phase flow variant of the well known PUNQ-S3 problem (Floris *et al.*, 2001). The complexity of this case study allows us for stressing some of the shortcomings of the traditional EnKF. Chapter 5 deals with the EnKF when applied to facies models.

The last group of chapters, Chapters 6 and 7, introduces newly developed algorithms combining the traditional EnKF with two parameterization methods capable of preserving the spatial variability of petrophysical properties. This feature makes it possible to ensure that the updated geological model is still consistent with the characteristics determined by geologists. Chapter 6 deals with the "Pilot Point Method" (de Marsily, 1984). After introducing this parameterization technique, we show how it can be combined with the EnKF. Then we apply this modified version of the EnKF to the same case study as the one presented in Chapter 3. Finally, we show how the proposed method could be improved. In Chapter 7, we consider "Gradual Deformation method" (Hu, 2000). After recapping the main outlines of this method,

we focus on its combination with the EnKF. The proposed algorithm is applied to the simple synthetic case study of Chapter 3.

2 State of the Art

In many fields, such as fluid flow in porous media, problems consist of identifying the state of a system that changes over time using a sequence of noisy measurements made on the system. The state of the system, ψ, involves all the variables required to describe the system under investigation. In order to analyze, draw conclusions and make forecasts about a dynamic system, at least two models are required: first, the “system model”, which is a model describing the evolution of the state with time and second, the “measurement model”, which is a model relating the noisy measurements to the state (Arulampalam *et al.*, 2002).

In most cases, states, measurements, initial and boundary conditions for the system are subject to inherent uncertainties or errors. Therefore, a probabilistic framework is suitable for any processing of the system. On the other hand, the processing of the system should correctly account for two sources of information:

- the a priori information, that is the initial knowledge about the system;
- the likelihood information, that is the information inferred from the measurements.

Identifying the state of a system from data is referred to as an "Inverse problem" (Tarantola, 1987). When this problem is solved, it provides us with "a posteriori" information on the system. The "Bayesian" approach is the cornerstone of the problem. It provides a rigorous general framework for dynamic state estimation problems, which also applies to fluid flow in porous media and, in particular, to the problem of history-matching.

In this chapter, we first introduce the Bayesian approach and a few methodologies developed to solve inverse problems. Section 2.2 is devoted to the Ensemble Kalman Filtering methodology, its historical motivations, assumptions and formulation. The Ensemble Kalman filter (EnKF) has been applied to several fields of science, such as meteorology, oceanography, hydrogeology and recently history-matching. Section 2.3 lists the advantages of the EnKF for history-matching and reviews its main applications in this specific field. Nevertheless, the EnKF has its own shortcomings and disadvantages: section 2.4 introduces the limitations imposed by the inherent assumptions of the EnKF and discusses several existing proposals in the literature for circumventing these problems. The discussion in sections 2.1, 2.2 and 2.4 is general while Section 2.3 addresses more specifically fluid flow in porous media and history-matching.

2.1 Bayesian Formulation

In the following discussion, we consider a dynamic system whose state is represented by ψ. The state of this system is composed of both static and dynamic variables. The first ones do not depend on time contrary to the last ones. In the remaining of this study, static variables may be referred to as parameters. The following derivations are based on the work of Evensen (2007).

It is assumed that an a priori information about the system is available at each time. The corresponding probability density function (pdf) is denoted by $f(\psi)$. At the initial time, $t = 0$, based upon preliminary model studies and given the initial and boundary conditions, this probability pdf is $f(\psi_0)$. Furthermore, it is assumed that over time, several measurements are performed, which can be directly or indirectly related to system variables. The sequence of these measurements is denoted by d and the associated likelihood pdf is denoted by $f(d|\psi)$. Referring to Bayes' theorem, the a posteriori pdf at each time can be written as

$$f(\psi|d) \propto f(\psi) f(d|\psi). \quad 2.1.$$

As a general rule, the dynamic model is discretized over time: the time intervals can be on a regular or irregular basis. At time t_k, the state of the system is ψ_k. The prior pdf at time t_k is

$$f(\psi) = f(\psi_0, \psi_1, \psi_2, ..., \psi_k) \quad 2.2.$$

On the other hand, the set of measurements collected over time is denoted

$$d = \{d_1, d_2, ..., d_N\} \quad 2.3.$$

It should be mentioned that the data acquisition times do not necessarily correspond to the times t_k associated to the ψ_k states: usually, their frequency is smaller.

There exist several solutions for this inverse problem. The rest of this section recaps briefly the main methods. The interested reader can refer to Seiler *et al.* (2009) for a thorough review. These methods can be grouped into "variational methods" and "sequential methods".

2.1.1 Variational Methods

In variational methods, one tries to acquire the best estimate of the model at a specified time, given some prior knowledge at the initial time, *i.e.* $t = 0$, and a set of measurements obtained at and before this time. Generally, the intermediate model states are not directly accounted for

in this formulation. The general workflow with variational methods consists of defining a cost function measuring the discrepancies between an estimated model and the actual geological formation. The objective is to obtain the best estimate that minimizes this cost function. There exist three main assumptions to define the mostly used form of the cost function for history-matching purposes. First, state variables and measurements are assumed to have a Gaussian distribution. Second, the dynamical model is assumed to be perfect, *i.e.* no model error exists. Third, the initial conditions for the dynamical model are known without uncertainty. Therefore, the cost function, $J(\psi)$, is defined as:

$$J(\psi) = (\psi - \psi_0)^t C_{\psi\psi}{}^{-1} (\psi - \psi_0) + (\psi - d)^t C_{\varepsilon\varepsilon}{}^{-1} (\psi - d) \qquad 2.4.$$

The first term on the right hand side evaluates the distance between the prior model, ψ_0, and the estimated model, ψ, whereas the second term measures the mismatch between the measurements, d, and the corresponding values inferred from the estimated model. Covariance matrix $C_{\psi\psi}$ characterizes the uncertainties in the prior model and covariance matrix $C_{\varepsilon\varepsilon}$ quantifies the experimental and theoretical uncertainties in the measured values. A best estimate should be searched for in a very large space. The main challenges are the highly nonlinear form of the cost function, the possible existence of high numbers of local minima and the identification of the area from which the best estimate must be searched for. The methods used to solve this optimization can be classified as "Gradient-based methods" and "Non gradient-based methods" (Le Ravalec-Dupin, 2005).

In gradient-based methods, the minimization of the cost function is driven by the gradients (or derivatives) of the cost function. Those ones can be computed using the adjoint-based methods (Chavent *et al.,* 1973) or approximated from finite-differences. A review and discussion about the existing methods can be found in Tarantola (1987), Press *et al.* (1988) and Sun (1994). It should be noted that gradient computation is very costly in terms of computation time. Gradient methods yield local optima. In addition, they provide only one single best estimate without uncertainties around it. To get a second constrained model, the whole optimization process must be repeated. Another issue is the Gaussian hypothesis. It would be of interest to see how it can be relaxed.

2.1.2 Sequential Methods

Another interpretation of Bayes' theorem (Evensen, 2007) consists of processing the system under consideration sequentially in time. It simply means that the solution for the system at

each time provides the prior model for the following time period. Moreover, at the end of this time period, the model is adjusted to regenerate measurements to be compared to the measurements collected during this period of time. Whatever the adjustment performed, the prior model must still be satisfied. The adjustment process in each time interval is an inverse problem by itself. It can be solved by any minimization method.

Sequential methods rely on two main assumptions. First, the dynamic model must be a first order Markov process, meaning that the state of the system at any time t_k, *i.e.* ψ_k, must depend only on the state of the system at previous time, t_{k-1}, *i.e.* ψ_{k-1}. This property is not a strict one: it is satisfied by most of the time-dependent systems. Because of this assumption, equation 2.2 is rewritten as:

$$f(\psi_0,\psi_1,\psi_2,...,\psi_k) \propto f(\psi_0)\prod_{i=1}^{k} f(\psi_i|\psi_{i-1}) \tag{2.5.}$$

The second assumption is that measurement data must be independent and uncorrelated in time. Therefore, the likelihood pdf for the time t_k is given as:

$$f(d|\psi) = \prod_{j=1}^{J} f(d_j|\psi_{i(j)}) \tag{2.6.}$$

J is the total number of measurements at and before time t_k and $\psi_{i(j)}$ corresponds to the model state at the time where measurement d_j is collected. Substituting 2.5 and 2.6 in 2.1 leads to the following form for the a posteriori pdf at time t_k:

$$f(\psi_0,\psi_1,...,\psi_k|d) \propto f(\psi_0)\prod_{i=1}^{k} f(\psi_i|\psi_{i-1})\prod_{j=1}^{J} f(d_j|\psi_{i(j)}) \tag{2.7.}$$

Now, according to Evensen (2007), we are capable of writing the above equation in a new order where the two products are expanded between times $t_{i(j)}, j=1,...,J$ when the measurements $d_j, j=1,...,J$ are obtained:

$$\begin{aligned} f(\psi_0,\psi_1,...,\psi_k|d) \propto\ & f(\psi_0) \\ & \prod_{i=1}^{i(1)} f(\psi_i|\psi_{i-1}) f(d_1|\psi_{i(1)}) \\ & \vdots \\ & \prod_{i=i(J-1)+1}^{i(J)} f(\psi_i|\psi_{i-1}) f(d_J|\psi_{i(J)}) \\ & \prod_{i=i(J)+1}^{k} f(\psi_i|\psi_{i-1}). \end{aligned} \tag{2.8.}$$

On the other hand, we can introduce measurements sequentially in time. Thus, we obtain the following formulation which is equivalent to the direct application of equation 2.8:

$$f(\psi_0,\psi_1,...,\psi_{i(1)}|d_1) \propto f(\psi_0) \prod_{i=1}^{i(1)} f(\psi_i|\psi_{i-1}) f(d_1|\psi_{i(1)}) \quad 2.9.$$

$$f(\psi_0,\psi_1,...,\psi_{i(2)}|d_1,d_2) \propto f(\psi_0,\psi_1,...,\psi_{i(1)}|d_1) \prod_{i=i(1)+1}^{i(2)} f(\psi_i|\psi_{i-1}) f(d_2|\psi_{i(2)}) \quad 2.10.$$

$$\vdots$$

$$f(\psi_0,\psi_1,...,\psi_{i(J)}|d_1,...,d_J) \propto f(\psi_0,\psi_1,...,\psi_{i(J-1)}|d_1,...,d_{J-1}) \prod_{i=i(J-1)+1}^{i(J)} f(\psi_i|\psi_{i-1}) f(d_J|\psi_{i(J)}) \quad 2.11.$$

$$f(\psi_0,\psi_1,...,\psi_k|d_1,...,d_J) \propto f(\psi_0,\psi_1,...,\psi_{i(J)}|d_1,...,d_J) \prod_{i=i(J)+1}^{k} f(\psi_i|\psi_{i-1}). \quad 2.12.$$

From 2.8, it is evident that we can first calculate the a posteriori distribution at time $t_{i(1)}$ when measurement d_1 is obtained. Then, this first a posteriori distribution becomes the a priori pdf of the following time interval $[t_{i(1)}, t_{i(2)}]$, $t_{i(2)}$ being the time when d_2 is acquired and so forth as shown in equation 2.10. This process is repeated until all measurements are processed, (equation 2.11). Then the last equation, that is 2.12, gives prediction at time t_k. Equations 2.9, 2.10, 2.11 and 2.12 are known as the recursive form of the Bayes' theorem.

Among the existing methods to solve the inverse problem at each time step, we will focus on two of them: the particle filter and the Ensemble Kalman filter (EnKF). Both of them require the knowledge of the prior model and its uncertainties at each time step.

The particle filter by Doucet *et al.* (2001) involves a large number of particles or model states to approximate the pdf for the prior solution. During each time step, each of the particles is propagated through time according to the system model or dynamic equations. In addition, this approach makes it possible to account for model errors. Whenever the measurements are available, a Bayesian update is performed by re-sampling. It supplies us with a set of

realizations or particles that represent the a posteriori pdf. The particle filter is the general method for solving inverse problems when referring to a sequential framework. However, it requires a large number of particles to get an accurate estimation and so far, it is only feasible when the number of uncertain parameters is low.

The Ensemble Kalman Filtering is based upon the same idea as the particle filter. However, in this case, the update is a simple linear update similar to the Kalman filter update. The assumption behind this linear update is that the prior pdf can be represented by its first two moments only, *i.e.* the mean and covariance. As a result, the a posteriori pdf is always a Gaussian distribution. More details are available in Evensen (1994, 2007).

2.2 Ensemble Kalman Filter formulation

The Ensemble Kalman Filter is an extension of the Kalman filter (Kalman, 1960) proposed for handling nonlinear problems. As a recursive filter, the Kalman filter is capable of efficiently estimating the state of a linear dynamic system on the basis of inaccurate (or noisy) measurements. It can be derived from Bayes' rule seeking either a maximum likelihood estimate (Lorenc, 1986) or a minimum variance estimate (within estimation theory, Cohn, 1997). It can also be identified as a recursive least squares problem weighted by the inverse of the error covariance matrices (Wunsch, 1996). In this section, we first analyze the formulation of Kalman filters. Then, we move to its variants for slightly and highly nonlinear systems, *i.e.* the Extended Kalman Filter (EKF) and the Ensemble Kalman Filter (EnKF), respectively.

2.2.1 Kalman Filter and Extended Kalman Filter

As mentioned in general for dynamic systems, the Kalman filter is based upon two models: the system model and the measurement model. In a Kalman filter, first the study time is divided into several consecutive intervals and then, within each interval, the model is processed according to two stages: first, "forecast" and second "update" or "analysis". In the following discussion, superscript f stands for forecast and a for analysis.

Assume that we are in the k^{th} time step. During forecast, the model is propagated forward in time. Its initial state, ψ_{k-1}^{a}, is the current estimate and its initial uncertainties are defined by its error covariance matrix $C_{\psi\psi}^{a,k-1}$. At the end of this time interval, the forecasted state is denoted

as ψ_k^f . Because of the linearity of the dynamic system, the forecasted model is derived from a simple formulation as below:

$$\psi_k^f = F\psi_{k-1}^a + q_{k-1} \qquad 2.13.$$

F is the linear system model operator and q_{k-1} is the model error at this time step. The uncertainties in the forecasted model, ψ_k^f , can be obtained as the forecast error covariance, $C_{\psi\psi}^{f,k}$:

$$C_{\psi\psi}^{f,k} = FC_{\psi\psi}^{a,k-1}F^T + C_{qq}^{k-1} \qquad 2.14.$$

C_{qq}^{k-1} is the covariance error matrix representing model errors. When the model is assumed to be perfect, variables related to the model error, q_{k-1} and C_{qq}^{k-1}, are neglected in equations 2.13 and 2.14. This is the case for fluid flow in porous media.

At the end of the k^{th} time interval, a set of measurements d_k is introduced; these measurements are linear functions of the true state, ψ_k^t. The associated noise is ε_k and the uncertainties in ε_k are given by the error covariance $C_{\varepsilon\varepsilon}$. The following equation provides the measurement model for this linear measurement system:

$$d_k = H\psi_k^t + \varepsilon_k \qquad 2.15.$$

H is the linear measurement model operator. During the update stage of a Kalman filter, the forecast estimate, ψ_k^f , is corrected to take the measurements into account. This updated state, ψ_k^a , is derived from equations 2.16 and 2.17:

$$\psi_k^a = \psi_k^f + K_k(d_k - H\psi_k^f) \qquad 2.16.$$

$$K_k = C_{\psi\psi}^{f,k}H^T(HC_{\psi\psi}^{f,k}H^T + C_{\varepsilon\varepsilon})^{-1} \qquad 2.17.$$

where K_k is the Kalman gain matrix. The uncertainties associated to the updated state are defined by the update error covariance matrix $C_{\psi\psi}^{k,a}$ as:

$$C_{\psi\psi}^{k,a} = (I - K_kH)C_{\psi\psi}^{k,f} \qquad 2.18.$$

The updated state ψ_k^a is used as the initial estimate for the following time interval and its associated uncertainties $C_{\psi\psi}^{k,a}$ are then the prior model uncertainties.

The Extended Kalman Filter (EKF) was first suggested for dealing with nonlinear system model operator, *i.e.* $\psi_{k+1}^f = F(\psi_k^a)$. In such a case, the evolution of the model in time can be

tracked easily. However, it is more difficult to estimate the uncertainties, which are represented by the error covariance matrix. The solution proposed by the EKF consists of expanding the error covariance function around the forecast model according to the Taylor formulation and to discard moments of third and higher orders. More details about the derivations can be found in numerous books about control theory such as Jazwinski (1970) and Gelb (1974). When the model is slightly nonlinear, the EKF can provide acceptable estimations of the model state. However, for highly nonlinear and large scale problems, two main issues arise: first, the computation and storage of the error covariance matrix become infeasible and second, as the high order terms in the Taylor expansion of the error covariance matrix are neglected, the error in the estimations may infinitely increase. Therefore, a better handling of nonlinearities is required. The Ensemble Kalman Filter (Evensen 1994, 2007) was proposed as an alternative.

2.2.2 Ensemble Kalman Filter

For a nonlinear and large-scale system, the evolution of the error covariance matrix with time can be partially deduced from the analysis of several realizations of the system under consideration as proposed by Evensen (1994). In this regard, Evensen proposed the Ensemble Kalman Filter (EnKF) as an ensemble based Monte Carlo method. Assuming that an infinite number of realizations are available for the forecast and analysis, the error covariance matrices are calculated as:

$$C^f = \overline{(\psi^f - \psi^t)(\psi^f - \psi^t)^T} \qquad 2.19.$$

$$C^a = \overline{(\psi^a - \psi^t)(\psi^a - \psi^t)^T}$$

The overbar denotes the expectation or the average value, ψ^t is the actual or true model at the considered time step. However, an infinite number of realizations is not a realistic assumption and in real conditions, we limit ourselves to an ensemble (or group) of N realizations. Moreover, the truth is never known. Thus, in many applications, the truth is replaced by the ensemble mean model denoted by $\overline{\psi}$. Therefore, in any application of the EnKF, at the initial time, an ensemble of N realizations for different static and dynamic variables is formed. Then, during forecast, each of the realizations is propagated through time according to the model operator F . Its corresponding uncertainty or error covariance matrix is calculated by taking an average over the finite ensemble of realizations as:

$$C^f \cong C_e^f = \overline{(\psi^f - \overline{\psi})(\psi^f - \overline{\psi})^T} \quad 2.20.$$

Then, each ensemble member is updated following the simple Kalman filter update equation (equation 2.16). We thus obtain an updated ensemble of realizations. Once more, the error covariance matrix for this updated ensemble is calculated as:

$$C^a \cong C_e^a = \overline{(\psi^a - \overline{\psi})(\psi^a - \overline{\psi})^T} \quad 2.21.$$

In the above formulas, subscript e stands for the finite size ensemble. The notions of time step n and state $\psi\psi$ are disregarded for simplicity. From the above formulation, the dimensionality or rank of the ensembles is clearly smaller than the error covariances C^f and C^a. Moreover, in comparison with the EKF, the full covariance matrix does not need to be stored. In addition, as the size of the ensemble increases, the error in the Monte Carlo sampling decreases by $1/\sqrt{N}$ and a better estimation of the error covariance matrix can be obtained.

Burger *et al.* (1998) noted that measurements at each time step should be treated as random variables in order to prevent the collapse of the ensemble members to one single realization. To do so, an ensemble of random perturbations, v_k, with a zero mean and a covariance equal to the measurement error covariance $C_{\varepsilon\varepsilon}$ has to be generated at each measurement time t_k. Therefore, the ensemble of perturbed measurements includes N members. Its i^{th} member is denoted by:

$$d_{i,k} = d_k + v_{i,k} \quad 2.22.$$

Then, each ensemble member $\psi_{i,k}$ must be updated considering measurement $d_{i,k}$ instead of d_k in equation 2.16.

2.2.3 Ensemble Kalman Filter implementation for history-matching

We introduce a few notations before developing the EnKF implementation in a history-matching context. The model under consideration is the reservoir model. The state of this model changes with time because of fluid displacements. Flow equations (Aziz and Settari, 1979) are solved using a reservoir fluid flow simulator. In this study, we ran the reservoir simulator PumaFlow™ from IFP. According to the governing flow equations, reservoir state ψ is characterized at each time step from two sets of variables populating every grid block of the numerical reservoir model: they are parameters or static variables such as porosity or

permeability, indicated by m, and state or dynamic variables such as pressure and saturation, denoted $g(m)$. Function g corresponds to the set of equations describing flow in porous media through time. In addition, through the life of a reservoir, several observations, d, are obtained such as producing well pressure, oil production rate, gas/oil ratio, *etc.* on bottom-hole or surface conditions.

The state vector for each ensemble member j at time step k is denoted by:

$$\psi_{k,j} = \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \end{bmatrix} \quad 2.23.$$

In addition, as the relation between measurements, parameters and dynamic variables is not linear, the state matrix is augmented with numerical simulation results through measurement operator, H_k. This matrix is trivial, with only 0 and 1 as shown below:

$$d_{sim}^{k,j} = \begin{bmatrix} 0 & 0 & I \end{bmatrix} * \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \\ d_{sim}^{k,j} \end{bmatrix} = H_k * \psi_{k,j} \quad 2.24.$$

I is the identity matrix. The new state matrix thus becomes:

$$\psi_{k,j} = \begin{bmatrix} m_{k,j} \\ g(m_{k,j}) \\ d_{sim}^{k,j} \end{bmatrix} \quad 2.25.$$

Assuming that the ensemble has a size of N, the state matrix, holding the forecast state vectors $\psi_{k,j}^f, j = 1,...,N$ at time step k, is defined by:

$$A_k = (\psi_{k,1}^f, \psi_{k,2}^f, ..., \psi_{k,N}^f) \quad 2.26.$$

The time index, k, is dropped in the rest of this section for simplicity. The ensemble mean model is obtained by taking an average over all ensemble members:

$$\overline{A} = A \times 1_N \quad 2.27.$$

where 1_N is an N by N matrix where each element is equal to $1/N$. The ensemble mean is stored in each column of $\overline{A}$. The ensemble perturbations, A', are defined by the difference between the state matrix and its mean:

$$A' = A - \overline{A} \quad 2.28.$$

Then, the forecast ensemble error covariance matrix is calculated according to:

$$C_e^f = \frac{A'(A')^T}{N-1} \quad 2.29.$$

In the following discussion, the ensemble holding the perturbed measurements is denoted by:

$$D = (d_1, d_2, ..., d_N) \quad 2.30.$$

The ensemble of random perturbations for measurements (see equation 2.22) is stored in the matrix:

$$\gamma = (v_1, v_2, ..., v_N) \quad 2.31.$$

from which we can construct the ensemble representation of the error covariance matrix

$$C_{\varepsilon\varepsilon} \cong C_{\varepsilon\varepsilon}^e = \frac{\gamma\gamma^T}{N-1} \quad 2.32.$$

The standard analysis equation 2.16 can now be expressed in terms of the ensemble covariance matrices as:

$$A^a = A + C_e^f H^T (HC_e^f H^T + C_{\varepsilon\varepsilon}^e)^{-1}(D - HA) \quad 2.33.$$

where H is the measurement operator introduced in 2.24. The above formula can be reformulated as:

$$A^a = A + A'A'^T H^T (HA'A'^T H^T + \gamma\gamma^T)^{-1}(D - HA) \quad 2.34.$$

In practical implementations, we assume the independence between measurements, hence, $C_{\varepsilon\varepsilon}^e$ is considered to be diagonal. Furthermore, it is assumed that $(HA'A'^T H^T + \gamma\gamma^T)$ is of full rank and invertible. Then, the basic procedure to calculate the update is as follows:

1. Construct HA' and $(D - HA)$
2. Construct $HA'A'^T H^T + \gamma\gamma^T$
3. Construct $(HA'A'^T H^T + \gamma\gamma^T)^{-1}(D - HA)$
4. Compute $A'A'^T H^T (HA'A'^T H^T + \gamma\gamma^T)^{-1}(D - HA)$
5. Finally update the state matrix $A^a = A + A'A'^T H^T (HA'A'^T H^T + \gamma\gamma^T)^{-1}(D - HA)$.

One of the main drawbacks of this algorithm is the possible singularity of the matrix to be inverted. More details are available in Evensen (2003).

2.3 Using the Ensemble Kalman Filter for petroleum applications

Since its introduction, the Ensemble Kalman Filter has been widely used in different fields of science to sequentially incorporate measurements collected at successive times into a dynamic

system. There are many examples in oceanography (Haugen and Evensen, 2002), meteorology (Evensen and van Leeuwen, 1996), hydrogeology (Margulis *et al.*, 2002) and petroleum engineering. In the following discussion, different applications of the EnKF in reservoir engineering developed during the past decade are outlined.

The first application of the EnKF in petroleum industry, to our knowledge, is based on a real time data interpretation in underbalanced and low-head drilling by Lorentzen *et al.* (2001). In their application, several flow variables (static and dynamic) of a two phase flow in the well were adjusted by the EnKF on receipt of the measurements such as downhole and choke pressures. The corrected parameters then served for a better control of the drilling operations. Their experiments, based both on synthetic and experimental data, were shown to be reliable in terms of predictions of drilling operations and motivated the authors to apply the EnKF for the estimation of reservoir parameters (Lorentzen *et al.*, 2003). Then, Nævdal *et al.* (2002) applied the EnKF to a two dimensional near-well reservoir model. They aimed to constrain permeability models. They showed that the EnKF provides acceptable parameter estimations and consequently, improved predictions. Following this initial work, Nævdal *et al.* (2003) applied the same methodology to two 2D, three phase models (a synthetic and a simplified field model), aiming at adjusting the permeability fields. The obtained results, in terms of production data matching and estimation of the true permeability field, were promising. However, the method was stated as novice in this area. Gu and Oliver (2005) applied the EnKF to the three dimensional model PUNQ-S3 to estimate the porosity and permeability fields, as well as pressure and saturations in each grid block. This synthetic field case will be described in more detail in Chapter 4. Gu and Oliver (2005) found the EnKF method more efficient than the variational history-matching methods in terms of computational overburden. Similarly, Gao *et al.* (2005) applied the EnKF and another method called Randomized Maximum Likelihood (RML) to the PUNQ-S3 model. The RML is a less expensive approximation for the Markov Chain Monte Carlo (McMc) method in sampling the a posteriori distributions (Oliver *et al.*, 1996). They found several similarities between the two methods, but concluded that the EnKF outperformed the RML due to simplicity of its implementation. Besides, the EnKF was applied by Liu and Oliver (2005) to estimate facies boundaries for porosity and permeability fields. As facies models are not continuous, direct application of the EnKF was not straightforward. Therefore, the Plurigaussian method with a truncation threshold was used to determine the location of the boundaries. The main difficulty in this application was to ensure that the geological models respect facies observations at wells. The proposed solution was to perform iterations during the update step of the EnKF

until the facies model respects the facies observed at wells. In addition, the performance of the EnKF was compared with that of the RML mentioned above. Again, the conclusion was that the EnKF is superior to variational methods regarding computational efficiency and ease of implementation. Wen and Chen (2005) introduced a modification in the EnKF workflow by implementing a confirming step. The confirming step consists in using the most recently updated static model parameters to rerun the simulations for the current time step and update the dynamic variables. This ensures that the static and dynamic variables are always physically consistent. Lorentzen *et al.* (2005) performed several tests on the PUNQ-S3 case to determine the effects of different initial ensembles as well as different initial distributions. The obtained results stressed the robustness of the EnKF method: the ten final different ensembles (assimilated from the same initial distribution) provided almost the same predictions and spread around the true cumulative oil production. Moreover, the final ensembles estimated from the two different initial distributions provided the same type of prediction, indicating that data assimilation in this case was successful in correcting possible uncertainties in the initial prior distribution. However, it was expected that the final ensembles converge to the same posterior distribution, regardless of the choice of the initial ensemble. This supposition was not satisfied neither within ensembles coming from the same initial distribution nor within ensembles coming from different initial distributions. The authors did not investigate the results thoroughly. However, they suggested that it may be due to the finite size of the ensembles. In addition, the EnKF has been used for assimilation of 4D seismic data in combination with production data. As these seismic data are given as the difference between two seismic surveys, a combination of the EnKF with the EnKS (Ensemble Kalman Smoother, Evensen, 2007) had to be used. Skjervheim *et al.* (2005) applied this methodology to two case studies: a synthetic case study and a real field case study in North Sea. In the synthetic case, the assimilation of 4D seismic data jointly with production data resulted in better estimations of the permeability field. For the real field data, the use of this combination yielded in different permeability fields compared to the case where only production data were considered. However, the match for the production data was then slightly improved. Zafari and Reynolds (2005) investigated the analogy between the RML and the EnKF. Introducing two mathematical toy problems, they showed that the Gaussian assumption in the EnKF penalizes this method when dealing with multi modal distributions. On the contrary, the RML method approximates these distributions better. In addition, Zafari and Reynolds (2005) showed that for a linear problem, the confirming step suggested by Wen and Chen (2005) is not compatible with the theoretical framework of the EnKF. Some authors have used the

EnKF methodology with streamline simulators. In this regard, Choe and Park (2006) suggested regeneration of the ensemble members when the estimated error covariance reaches one fifth of its initial value. Furthermore, in their case study, they showed that by selecting the measurement data to which the static variables are more sensitive, the final porosity and permeability models can be better constrained. Another application of the EnKF with the streamline simulator is reported in the work of Arroyo-Negrete (2006). He claimed that some problems regarding the overestimation of parameters such as porosity and the non Gaussian distributions can be solved by restricting the effect of each measurement on the neighboring grid cells. He suggested that the covariance localization deduced from the streamlines is a potential solution for the mentioned problem.

Haugen *et al.* (2006) presented a successful application of the EnKF to a North Sea real field case using real production data. They proposed to include some other uncertain parameters such as fault transmissibilities, gas oil contact and water oil contact within the state matrix of the EnKF. Later, Evensen *et al.* (2007) studied a North Sea case, considering the parameters suggested by Haugen *et al.* (2006). Besides, they discussed the advantages of the EnKF over some other history-matching techniques. Liang *et al.* (2007) applied Singular Evolutive Interpolated Kalman Filter (SEIKF) to a three dimensional water flooding case where reservoir permeability was estimated by matching the production data. This filter operates in three steps: resampling, forecast and analysis. Unlike the EnKF, the SEIKF does not make direct use of the outcome of the previous assimilation step. Instead, the resampling step generates perturbations of the previous assimilation results to locally widen the space explored by the filter. They concluded that the SEIKF exhibits rapid convergence in the domain of the model parameters. In terms of accuracy and uncertainty reduction, the SEIKF performs almost the same as the EnKF. As will be detailed in section 2.4, the distant observations may have an unfavorable impact on some model parameter updates and one general solution is the covariance localization. In this regard, Devegowda *et al.* (2007) proposed a covariance localization method based upon streamline-based sensitivities to modify the cross covariance matrix to reduce the unwanted influences. This method circumvented many of the traditional EnKF problems such as instabilities, parameters overestimation and loss of geological realism. Gu and Oliver (2007) claimed that neither the Extended Kalman Filter (EKF) nor the Ensemble Kalman Filter (EnKF) are suitable for highly nonlinear problems such as those involving multiphase flow in porous media. They introduced Gauss-Newton iterations to improve the analysis step. They studied several cases and concluded that this new algorithm is superior to the EnKF to determine uncertainty.

Bianco *et al.* (2007) applied the EnKF to the Zagor field in West Africa. They aimed at estimating porosity fields from the collected production data. It was observed that the EnKF based assimilation process narrowed the production forecast confidence intervals with respect to the estimates derived from the prior distributions. Jafarpour and McLaughlin (2007) combined the EnKF with a parameterization technique, the Discrete Cosine Transform (DCT). The idea was to reduce the dimensions of the state vector through DCT parameterization, making the filtering approach much more practical for realistic large scale applications. These authors performed several numerical experiments and found that the results obtained with the DCT based EnKF method are almost identical to the ones obtained with the simple EnKF scheme. Li and Reynolds (2007) proposed three iterative approaches for the EnKF. They concluded that whether or not an iterative approach has to be used depends on the user definition of a sound history-matching. Wang *et al.* (2007) introduced the EnKF as an optimization algorithm, but concluded that, compared to other optimization algorithms, the EnKF requires significantly more CPU-time and yields poorer estimates. In addition, they proposed a closed-loop reservoir management with the EnKF for data assimilation and the steepest ascent for production optimization based on a central forecast model. It was observed that the results are reasonable for a small test example. Chen *et al.* (2008) introduced an ensemble based closed-loop optimization method. The new method combines a new ensemble-based optimization scheme (EnOpt) with the EnKF. It was observed that this method is able to regenerate the main geological features as well as to increase the net present value to a level comparable with the hypothetical case of optimizing based on known geology.

2.4 Ensemble Kalman Filter issues and solutions

The previous section presented several applications of the EnKF method. Most of them provided promising results. In addition, this technique is attractive because of its simplicity and computational efficiency. However, there are still some issues that were pointed out when using the EnKF as mentioned in the previous section. The main ones are related to the finite size of the ensemble, the Gaussian distribution assumption for the static and dynamic variables and the nonlinearity of the dependence between measurements and state variables. In the following subsections, an overview of each of these issues is provided along with the proposed solutions.

2.4.1 Finite ensemble size

The EnKF methodology is based upon a proper estimation of the error covariance matrix. This is the main reason why ensembles are considered. If the size of the ensemble is large enough, the error covariance matrix calculated over the ensemble provides us with the correct error covariance matrix. However, in real applications, the size of the ensemble is by far smaller than the state matrix. Therefore, the estimates of the covariance matrix can be biased. This phenomenon is often referred as "spurious correlations". These spurious correlations induce high correlations between distant points which are clearly undesirable. For distant points, correlations should be negligible. In practice, this problem shows itself as the strong effect of a measurement at a point or grid cell placed far from it. In addition, the rank of the error covariance matrix deduced from the ensemble is less than or equal to the size of the ensemble. The assimilation process being performed step by step, some degrees of freedom are successively removed. This limits the number of measurements that can be assimilated. This difficulty can be strongly penalizing when dealing with 4D seismic data. The two issues listed above can be at least partially solved by considering a larger ensemble. However, this is just not feasible in practical applications. Therefore, alternative methods have been proposed. They are generally called "Localization" methods. They involve either a neighborhood distance to limit the correlation values or a virtual increase of the size of the ensemble. These methods are discussed in the following subsections.

2.4.1.1 Distance-based localization methods

The most straightforward method to perform localization is to set the correlation between two points to zero when the distance between these two points exceeds a certain influence distance (Houtekamer and Mitchell, 1998). This idea was further developed by Houtekamer and Mitchell (2001). They first propose to perform localization by multiplying the error covariance matrix C by a weighting matrix ρ. Thus, they obtained the localized covariance matrix $C_{Loc} = \rho \circ C$ where $\circ$ is the matrix elementwise multiplication. The purpose of this product is to remove the undesired correlation values. Second, Houtekamer and Mitchell (2001) proposed to group the production data at each step into batches and to perform the EnKF update step based upon these batches instead of the individual data. The observations in each batch have correlated error covariances. The authors (2001) concluded that this algorithm results in smoother updated fields and that the so modified EnKF performs

accurately with a fairly small ensemble. Other examples stressing the usefulness of this approach can be found in Hamill *et al.* (2001). These authors also investigated the covariance inflation phenomenon, which allows the ensemble members to depart from the ensemble mean. They found that localization and inflation have a dual effect on the representation of the covariance matrix in terms of rank issues. Lorenc (2003) showed that when using the Gaspari Cohn function (Gaspari and Cohn, 1999) to compute the weight matrix, the localization radius increases with the ensemble size. As a result, very large ensembles do not suffer from spurious correlation effects.

An interesting idea was proposed by Furrer and Bengtsson (2007). They suggested to identify an optimal correlation function ρ_{opt} among a set of valid correlation functions through a minimization problem. Therefore, the selected correlation function depends on the true error covariance matrix type and the size of the ensemble.

2.4.1.2 Other localization methods

The state matrix can be composed of variables or parameters such as fault transmissibility and water-oil contact which are not distance dependent. Therefore, if localization is to be applied, another technique, which does not depend on distance, must be envisioned. One solution was proposed by Furrer and Bengtsson (2007): it consists of determining the elements of the correlation matrix with respect to the elements of the error covariance matrix itself. These types of correlation functions or matrices do not necessarily decrease monotonically with distance. Another type of distance-independent localization filter or function was proposed by Anderson (2007).

Another issue related to distances is the influence area (or radius) around a measurement. Indeed, this radius can evolve with time (Hacker *et al.*, 2007). Therefore, Bishop and Hodyss (2007) proposed an "adaptive" localization correlation function which is a function of both time and distance.

In special cases where sensitivity studies can determine the influence of a certain number of measurements on a certain number of state variables or parameters, more judicious localization methods can be developed. One example, as mentioned briefly in section 2.3, involves the estimation of sensitivities from streamline-based flow simulations. This allows for quantifying influence areas around measurements. Once sensitivities are known, they can be used to estimate cross-correlations between state variables and measurements (Arroyo-Negrete *et al.*, 2006; Devegowda, 2007). The method was shown to work well for a synthetic

case study but its potential for a real case, with large correlation lengths, has still to be discussed.

2.4.1.3 Local analysis

The state space in most data assimilation applications is of a very large size and it is difficult to capture all of the features of this large region from an ensemble of models with finite size. However, it can be assumed that the whole region can be divided into several smaller (or local) regions. Then, the size of the finite ensemble can be of the same order as that of the local region.

Keeping this idea in mind, Ott *et al.* (2004) proposed to modify the standard EnKF workflow as follows. Within each time step, forecast is performed considering all ensemble members. However, the analysis phase is performed after dividing the global state space into sub-regions. The update of the standard EnKF is then performed for each of these sub-regions independently. When done for all the local regions, a global analysis ensemble is formed. Last, this global ensemble is used as the initial ensemble of the following forecast step. The same idea is also explained in Evensen (2007).

Local analysis has the nice property of increasing virtually the size of the ensemble. Hence, it provides more degrees of freedom for dealing with large state spaces. However, the updated models can depart very significantly from the prior pdf, especially when the correlation lengths characterizing the porosity/permeability properties are much larger than the sizes of the sub-regions.

2.4.2 Non-Gaussianity and nonlinearity

Gaussianity of parameters and states is a fundamental assumption in all Kalman filters, including the EnKF. In reality, nature is barely Gaussian. For instance, permeability distributions are often considered as log-normal. Similarly, water saturations in water flooding applications do not respect any Gaussian distributions. In addition, even if the initial distributions are normal, the non-linearity of the dynamic model, *i.e.*, the fluid flow equations, may result in non-Gaussian distributions (Chen *et al.*, 2009). Moreover, the EnKF calls for an approximately linear relation between model variables (static and dynamic) and measurements, which is not usually satisfied. In this section we aim to address some solutions to relax these two assumptions, at least slightly, about Gaussianity and linearity.

The problem of non-Gaussianity can be dealt with a "parameterization" method to map non-Gaussian variables to a Gaussian one, that is the one updated through the EnKF methodology. Then, a backward transform makes it possible to come back to the non-Gaussian space. There exist various methods and applications in the literature, among which is the log-normal transform developed by Bertino *et al.* (2003) for phytoplankton concentrations. It was shown that this transform improves the performance of the filter. A normal score transform was also applied to water saturations by Gu and Oliver (2006). It prevented saturations from going out of a physical range, but not from oscillating between low and high values. Dealing with facies models boils down basically to the same type of difficulties. They will be discussed in detail in Chapter 5. Briefly, the techniques proposed to work with facies models involve an intermediate parameterization technique that is combined with the EnKF. Among the proposed methods, the Truncated Pluri-Gaussian (Liu and Oliver, 2005), Level Set method (Moreno and Aanonsen, 2007) and Gaussian Mixture methods (Dovera and Della Rossa, 2007) can be mentioned. The parameterization technique is chosen so that the dependence between the state variables and the measurements becomes more linear.

Other reparameterization methods were introduced by Gu and Oliver (2006) and Chen *et al.* (2009). These authors substituted water saturations in the state matrix by other parameters depending on the water front location in the first work and on the water arrival time in the second work.

Another solution involves iterative approaches, iterations being performed to ensure that the updated values are still physical. Some of these iterative algorithms are explained and applied on petroleum case studies by Reynolds *et al.* (2006), Li and Reynolds (2007) and Gu and Oliver (2007). In iterative methods, the usual principal aspect is the definition of an objective function for each realization within the ensemble. Those are minimized after each forecast. Such methods depend on the optimization method selected for minimizing the objective function. They also differ based upon the starting time from which the forecast model is run: it can be zero or the previous assimilation time. More details can be found in the thorough review by Aanonsen *et al.* (2009).

2.5 Conclusions

This chapter has introduced the main techniques presented in the literature for solving inverse problems and more especially history-matching. We focused on the Kalman Filter family,

which includes the Kalman Filter (KF) for linear dynamic models, the Extended Kalman Filter (EKF) for slightly nonlinear models, and the Ensemble Kalman Filter (EnKF) for highly nonlinear models.

Fluid flow in porous media is a nonlinear model. Besides, it involves a huge amount of parameters. Therefore, the EnKF can be considered as an appropriate sequential method for handling history-matching. We provided the mathematical background for the EnKF and recapped the main works this method motivated during the past decade. We attempted to identify the main advantages and disadvantages of the EnKF. This chapter is the backbone for the research performed during this thesis and is considered as the reference in the following chapters.

I. Ensemble Kalman Filter Application

3 Two Dimensional Synthetic Case Study

In this chapter, we aim to apply the EnKF method to a 2D synthetic water flooding case study to assess the performance of the method in history-matching and production prediction. In section 3.1, we first present the case study and the EnKF settings. The method involves several parameters, which can control the performance of the filter. These parameters are the ensemble size, the assimilation time step, the measurement uncertainty level and the choice of the initial ensemble. Their influence was investigated with various sensitivity tests as shown in Section 3.2. These experiments provide guidelines on how they must be selected.

3.1 History-matching of the reference case

We study a very simple case so that we can analyze various aspects of history-matching with the EnKF. This case study consists of a two dimensional, two phase flow (water and oil), water flooding synthetic reservoir. The only uncertain parameter is permeability. The other petrophysical parameters such as porosity, relative permeability curves, *etc.*, are assumed to be known. The production data are the bottom hole pressure, surface oil rate and water cut. The water cut is the ratio of volume of water produced to the volume of all liquids produced. These data are known with a given level of noise or uncertainty. A preliminary step consists of generating the reference permeability model and simulating the bottom hole pressure, oil rate and water cut over 17 months. The aim of the experiment is then to regenerate the reference permeability field by performing a history-match (or assimilation) of these production data at the end of each month.

3.1.1 2D case study and the EnKF scheme

The numerical model for the 2D case study is built over a 50×50 grid. Dimensions of grid blocks are 10m along axes X and Y. The Fast-Fourier Transform-Moving Average (FFT-MA) algorithm (Le Ravalec *et al.*, 2000) was used to generate a permeability field k and populate the grid. After logarithmic transformation, the $\ln(k)$ field has a Gaussian histogram with a mean of 5.0 and a variance of 2.5. Its variogram is an anisotropical spherical variogram with

the main anisotropy direction defined by an angle of 30 degrees with respect to the X-axis. The horizontal variogram ranges are 200m and 40m. The reference log- permeability field is shown in Figure 3-1.

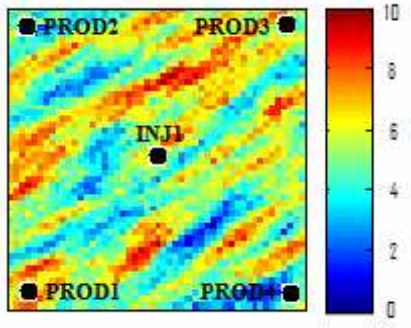
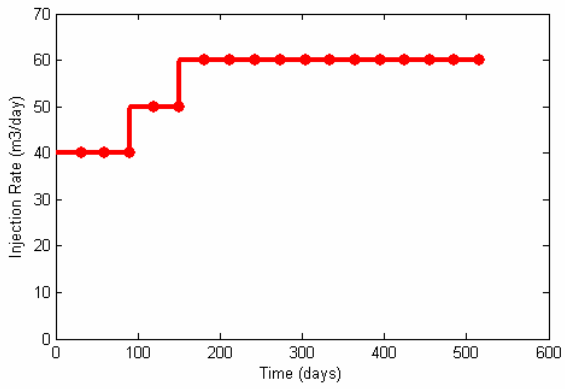


Figure 3-1 Reference ln k field

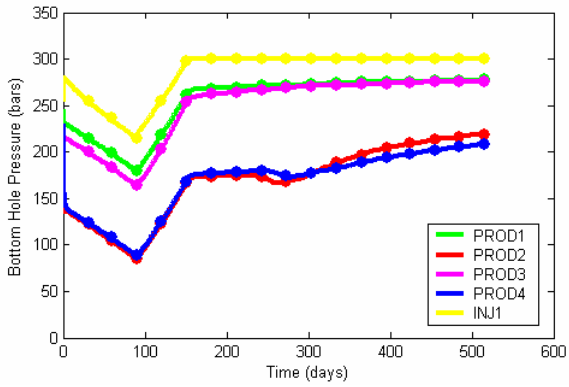
Five vertical wells are located in the field: one injector (INJ) in the center and four producers (PROD1 to PROD4, clockwise) in the four corners (Figure 3-1). The top depth of the reservoir is 2200m and the initial reservoir pressure is 250 bars. All producers produce with a rate of 10 m^3/day. The flow rate constraint is changed to a bottom hole pressure (BHP) control when BHP drops below 50 bars. The primary constraint for the injection well is the maximum injection rate, initially set to 40 m^3/day. It is then changed to 50 m^3/day and 60 m^3/day after 90 and 151 days of production, respectively (Figure 3-2 (a)). The secondary constraint for this well is the maximum injection pressure of 300 bars. The STONE1 model has been used for defining relative permeability curves with the associated end points and water capillary pressure tables. The initial water saturation is set to 0.15 and the residual oil saturation to 0.2. The compressibilities of water and oil are set to 0.10×10^{-4} bar^{-1} and 0.10×10^{-3} bar^{-1}, respectively. The corresponding densities are 1.05 g/cm^3 and 0.9 g/cm^3. The industrial IFP reservoir simulator PumaFlow™ is used for reservoir simulation, running from day 0 up to day 516 with a Black Oil PVT type. The bottom hole pressure (BHP), surface oil rate (SOR) and water cut (WCT) in all wells, computed for the reference permeability model, are shown in Figure 3-2 (b), Figure 3-2 (c) and Figure 3-2 (d). They are now identified as the reference production data.

The reference permeability field (Figure 3-1) is the model we would like to regenerate by assimilating the reference production data. At this point, we consider that we do not know the reference permeability field while we know the production data. Thus, we use the EnKF to identify a set of realizations conditioned to these production data. The production data are assimilated at the end of each month up to 516 days (17 months). The assimilation times are represented by solid dots in Figure 3-2 and are indicated below:

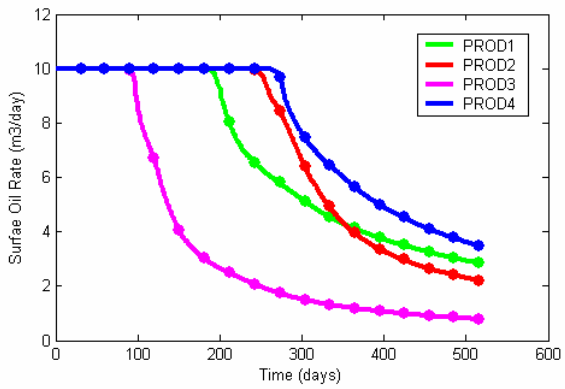
$t_{assi} = [31,59,90,120,151,181,212,243,273,304,334,365,396,424,455,485,516]$



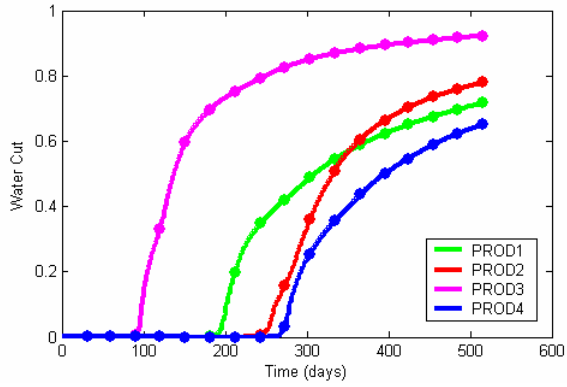
(a) Maximum injection rate planning



(b) Bottom hole pressure for the reference case



(c) Surface oil rate for the reference case



(d) Water cut for the reference case

Figure 3-2 (a) Injection rate (b) Bottom hole pressures (c) Surface oil rates (d) Water cuts. Curves in plots b to d are computed for the reference permeability model. The solid dots indicate the timing at which production data will be assimilated with the EnKF.

The distributions of measurement errors are assumed to be Gaussian with standard deviations of 5% of the values measured for BHP, SOR and WCT.

An initial ensemble of 50 permeability models is generated using the FFT-MA method with the same histogram and variogram as the reference field. Other parameters (porosity = 0.18, relative permeability curves, initial pressure = 250 bars and initial water saturation = 0.15) are kept constant without any uncertainty.

The state vector in this case study for each ensemble member j is composed of:

- The logarithm of the permeability values at each grid cell (static variable): $\ln k_{j,i}$, $i = 1,...,2500$;
- Pressure at all grid cells (dynamic variable): $P_{j,i}$, $i = 1,...,2500$;
- Water saturation at all grid cells (dynamic variable): $Sw_{j,i}$, $i = 1,...,2500$;
- Bottom hole pressure at injection and production wells: $BHP_{j,k}$, $k = 1,...,5$;

- Surface oil rate at production wells: $SOR_{j,k}$, $k = 1,...,4$;
- Water cut at production wells: $WCT_{j,k}$, $k = 1,...,4$.

Therefore, the state vector for the j[th] ensemble member is defined by:

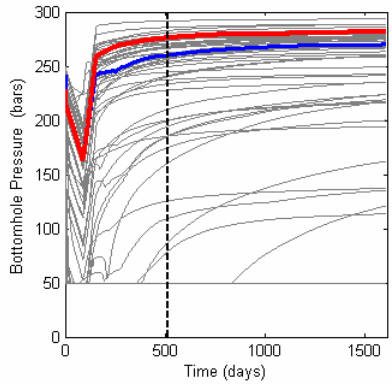
$$\psi_j = [\ln k_{j,1}, \ln k_{j,2}, ..., \ln k_{j,2500}, P_{j,1}, P_{j,2}, ..., P_{j,2500}, Sw_{j,1}, Sw_{j,2}, ..., Sw_{j,2500}, \\ BHP_{j,1}, ..., BHP_{j,5}, SOR_{j,1}, ..., SOR_{j,4}, WCT_{j,1}, ..., WCT_{j,4}] \quad (3.1.)$$

After performing the assimilations, the study time is expanded from 516 to 1612 days to assess the prediction capability of the updated (or corrected) permeability fields.

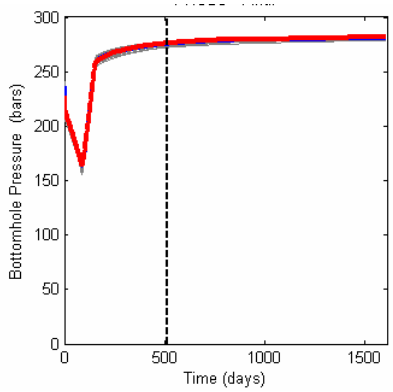
3.1.2 History-matched production data

In this section, we compare the performance of the initial and corrected models for two of the production wells: PROD3 and PROD4. PROD3 is located in a high permeability region (top right corner). It is connected to the injector through a high permeability streak (Figure 3-1). As expected, the earliest water breakthrough is observed at this well. On the contrary, PROD4 is located in a low permeability region (bottom right corner of Figure 3-1). This well has the latest water break through time. It produces oil at the required flow rate longer than the others. The production responses obtained for the initial and final models are shown in Figure 3-3 and Figure 3-4. For both figures, first column displays the production responses simulated for the members of the initial ensemble. The figures in the second column have been obtained performing a reservoir simulation over the whole time span with the final permeability fields taken back to day 0. In addition, the ensemble member responses (gray) are compared to the reference model responses (red). The blue curve corresponds to the mean permeability model: it was obtained by averaging the initial and final ensembles. The black dashed line splits the time axis into two phases: the history-matching phase (0 to 516 days) and the prediction phase (517 to 1612 days).

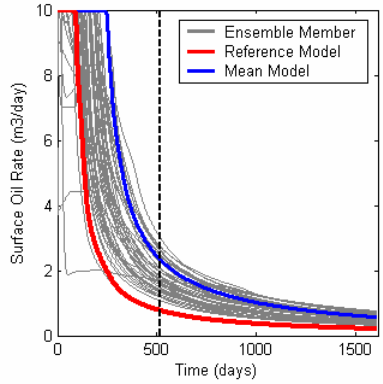
For well PROD3, we observe that initial models show greater variation in performance than corrected models. Figure 3-3(a)-(c) show that some of the initial models can not keep track of the reference well production data. After correction, all models respect the required targets (Figure 3-3(d)). Even in the prediction phase, the corrected models produce according to the desired scenario and perform as the reference case.



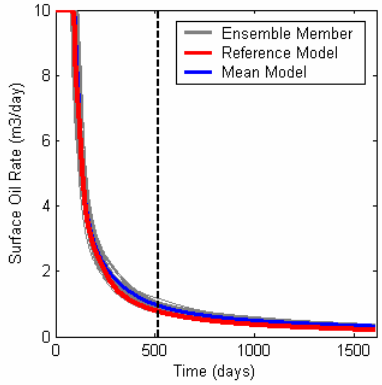
(a) Bottom hole pressure in PROD3 – Initial Ensemble



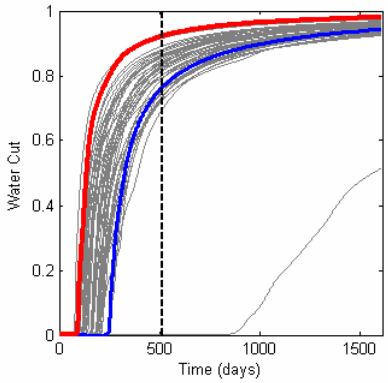
(d) Bottom hole pressure in PROD3 – Final Ensemble



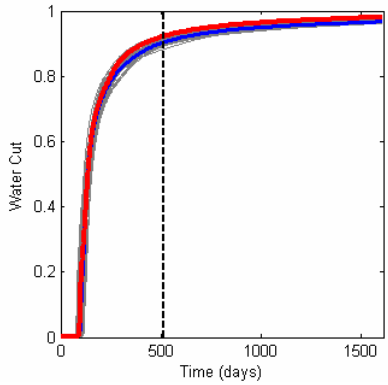
(b) Surface oil rate in PROD3 – Initial Ensemble



(e) Surface oil rate in PROD3 – Final Ensemble



(c) Water cut in PROD3 – Initial Ensemble

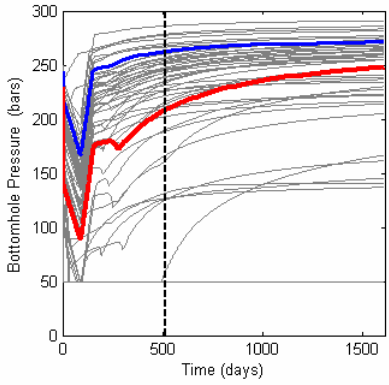


(f) Water cut in PROD3 – Final Ensemble

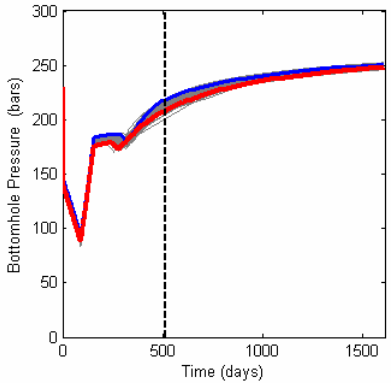
Figure 3-3 Production data simulated at well PROD3 with the 50 permeability realizations of the initial ensemble (first column) and final ensemble (second column). The black dashed line at 516 days indicates the time limit between the matching period (0-516 days) and the prediction one (517-1612 days). The gray curves correspond to the ensemble members and the blue one to the ensemble mean model. The reference data are plotted in red.

Comparison of Figure 3-3(b) and Figure 3-3(e) shows substantial improvements in the fit of surface oil rate. Most of the initial models produce more oil than the true model does while the corrected models have a surface oil rate close to the reference one. Figure 3-3(c) and Figure 3-3(f) show that the water breakthrough times computed for the final models are grouped around the reference while they show significant dispersion for the initial models. Data assimilation thus provides an ensemble of permeability values for which the simulated

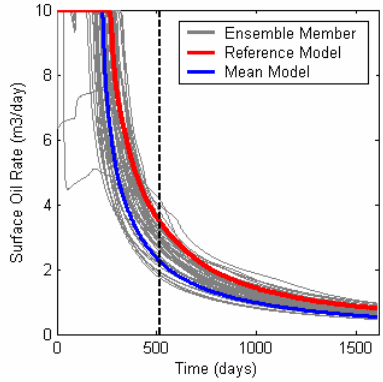
production responses are much closer to the reference ones, with a reduced spread compared to the initial ensemble.



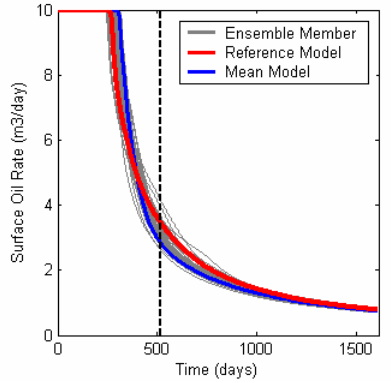
(a) Bottom hole pressure in PROD4 – Initial Ensemble



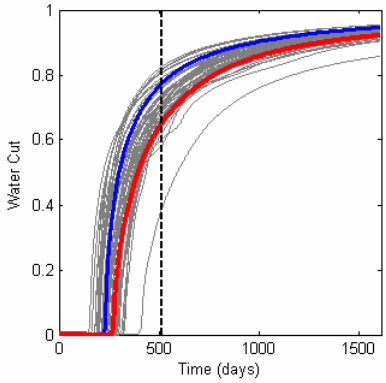
(d) Bottom hole pressure in PROD4 – Final Ensemble



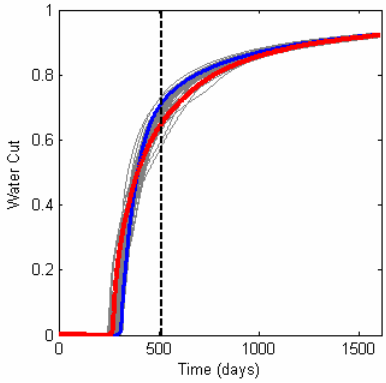
(b) Surface oil rate in PROD4 – Initial Ensemble



(e) Surface oil rate in PROD4 – Final Ensemble



(c) Water cut in PROD4 – Initial Ensemble



(f) Water cut in PROD4 – Final Ensemble

Figure 3-4 Production data simulated at well PROD4 with the 50 permeability realizations of the initial ensemble (first column) and final ensemble (second column). The black dashed line at 516 days indicates the time limit between the matching period (0-516 days) and the prediction one (517-1612 days). The gray curves correspond to the ensemble members and the blue one to the ensemble mean model. Reference data are plotted in red.

As can be seen in Figure 3-4, the same kind of results is obtained for well PROD4. The variation within the ensemble of corrected models is still reduced. The corrected permeability models allow us to estimate the uncertainty in future productions. Variations in the production

responses computed for the updated ensemble is larger for well PROD4 than for well PROD3. This result may be related to the water breakthrough which arrives later. For both wells, the initial uncertainty was reduced further for the BHP than for SOR and WCT.

An indicator widely used to quantify the quality of the updated model at each assimilation time step is the Root Mean Square (RMS) error for the well production responses. There are two forms of it. The first one assesses the departure of production responses for each ensemble member from the reference ones. The second error considers the production responses determined with a mean model given by the average over the ensemble. In both cases, the updated ensemble models are taken back to time zero and simulated over the history-matching period (0-516 days). We consider these two RMS errors.

The first one is referred to as $RMS_{prod,1}$. The subscript $prod$ indicates that the RMS error is calculated over production responses.

$$RMS_{prod,1} = \sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}\frac{1}{N_t}\sum_{j=1}^{N_t}\frac{1}{N_{data}^{t_j}}\sum_{k=1}^{N_{data}^{t_j}}(d_{obs}^{k}(t_j) - d_i^{k}(t_j))^2} \qquad 3.2.$$

N_e is the ensemble size, N_t the number of data assimilation time steps, $N_{data}^{t_j}$ the number of data collected at each time step (as a general rule, it is determined by the number of wells at which the production data of interest is observed). i is the ensemble member index, j the time index (time is denoted by t_j) and k the production response index. $d_{obs}^{k}(t_j)$ is the k^{th} data of the reference model at time t_j and $d_i^{k}(t_j)$ is the k^{th} data simulated for the i^{th} ensemble member at time t_j. Two $RMS_{prod,1}$ values are computed here, considering production responses simulated from time 0 with both initial and final ensemble members. They are reported in Table 3-1. Thanks to data assimilation, the RMS errors for the BHP, SOR and WCT are decreased to 8%, 23% and 24% of their initial values. The bottom hole pressure match is better than the SOR and WCT matches. One possible explanation is the relation between state variables, *i.e.* permeability fields, and numerical responses: it may be more linear for BHP than for SOR and WCT.

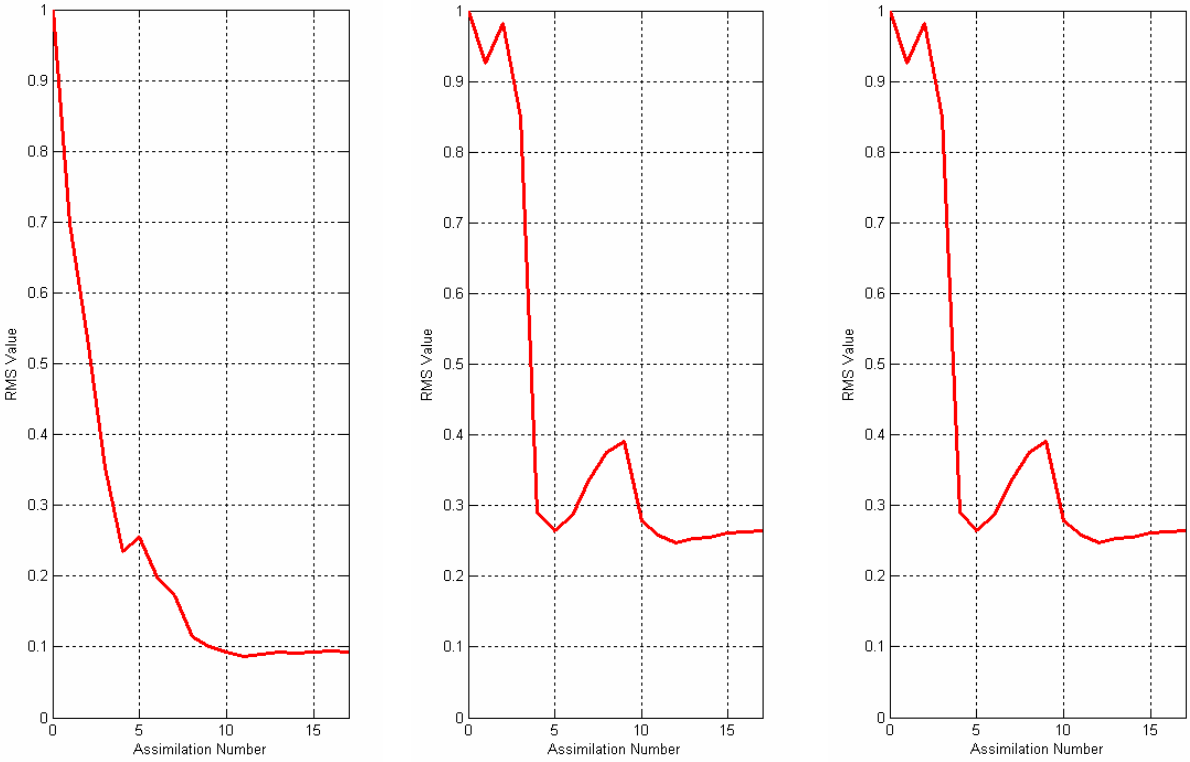
Table 3-1 $RMS_{prod,1}$ for the initial and final ensembles.

$RMS_{prod,1}$	BHP	SOR	WCT
Initial	7.2302	0.3077	0.0304
Final	0.6042	0.0738	0.0074
Final / Initial	0.0836	0.2398	0.2434

The second RMS error introduced above is referred to as $RMS_{prod,2}$. After each assimilation time step, a fluid flow simulation is performed for the mean updated permeability model from time zero to the current time step. Then, the RMS error is computed only over this period by:

$$RMS_{prod,2}(l) = \sqrt{\frac{1}{N_t^l}\sum_{j=1}^{N_t^l}\frac{1}{N_{data}^{t_j}}\sum_{k=1}^{N_{data}^{t_j}}(d_{obs}^k(t_j) - d_{mean,l}^k(t_j))^2} \quad 3.3.$$

Notations are the same as the ones introduced above for equation 3.2. In addition, l is the assimilation time index and N_t^l the number of assimilation times up to and including time step l. $d_{mean,l}^k(t_j)$ is the k^{th} data simulated at time t_j for the mean reservoir model obtained after assimilation at time t_l. In the rest of this chapter, these values are presented while normalized to the initial RMS value.



(a) Bottom hole pressure (bars) (b) Surface oil rate m3/day (c) Water cut

Figure 3-5 $RMS_{prod,2}$ computed for (a) Bottom hole pressure, (b) Surface oil rate, (c) Water cut throughout the assimilation procedure.

As can be seen in Figure 3-5, the $RMS_{prod,2}$ values for BHP, SOR and WCT follow a general decreasing trend as the assimilation process goes on. A strong decrease is observed between the 3rd and 4th assimilation steps for the SOR and WCT. It corresponds to the period of time during which there is the water breakthrough and the oil production rate decline at well PROD3. There is also a slight RMS increase for the SOR and WCT after the 6th assimilation

step. At this time, the injection rate is increased which results in strong non linear changes in the dynamic behavior. In addition, during the few following time steps, water breakthrough happens at wells PROD2 and PROD4. This can also impact the RMS behavior. It seems that the EnKF process needs a few additional time steps to assimilate the changes in the flow regime. After a few assimilation steps, the RMS comes back to the expected decreasing behavior.

3.1.3 History-matched permeability model

Another important issue in any history-matching problem is the ability of the method to approach the reference permeability field. As mentioned previously, the mean of the $\ln k$ field is 5.0 and its variance is 2.5. Figure 3-6(a) shows the true $\ln k$ field. For comparison purposes, the evolution of the ensemble mean along the assimilation procedure is displayed in Figure 3-7(b)-(s).

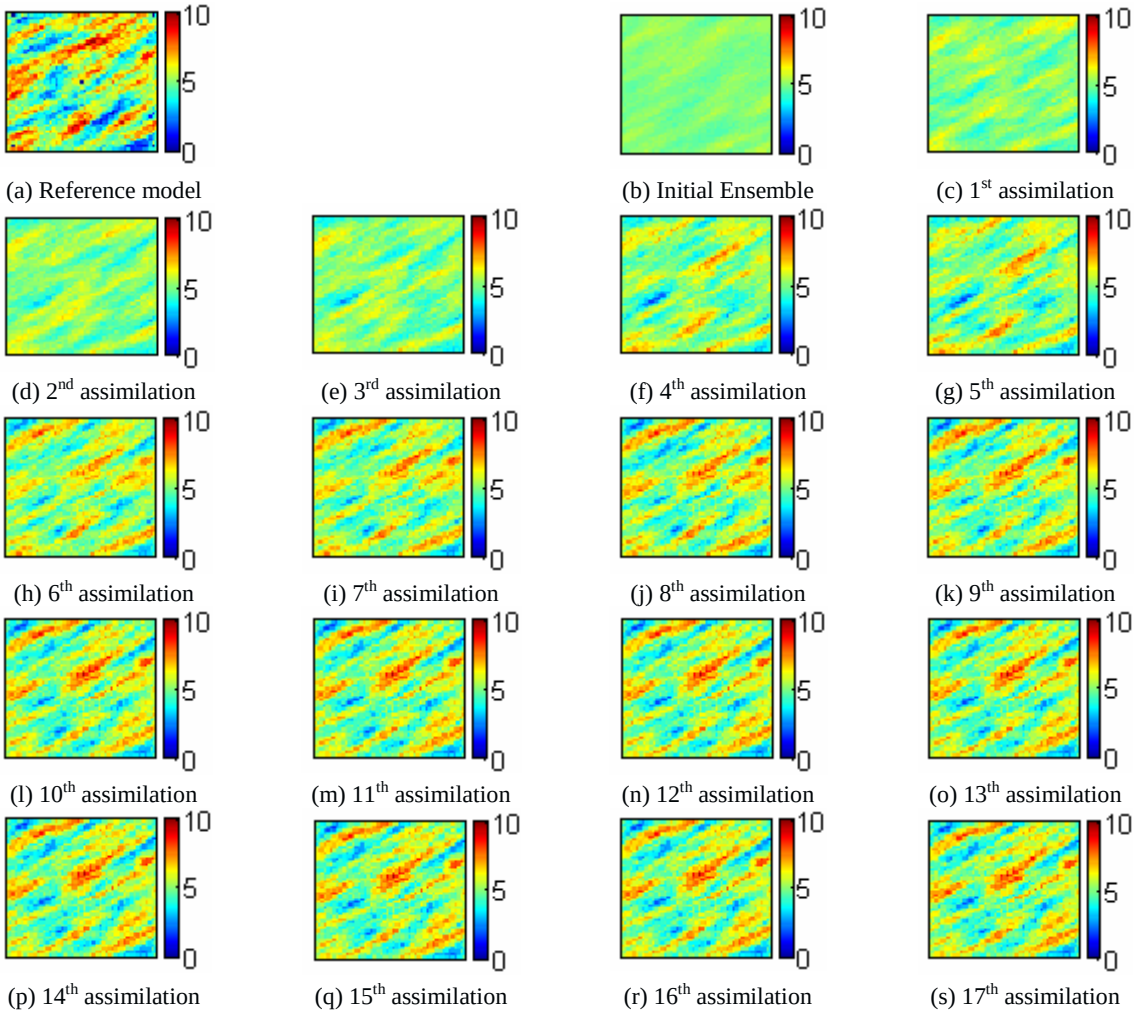


Figure 3-6 ln k evolution through time: (a) Reference Model, (b) Initial Mean Model, (c)-(s) Mean Model after assimilation 1-17.

The true $\ln k$ field is characterized by low- and high-permeability bands. The mean of the initial $\ln k$ ensemble is almost uniform and equal to the stationary mean 5.0. The assimilation process contributes to stress heterogeneities. After the assimilation corresponding to the 4th month, we start to see the banded features (this assimilation step corresponds to the first oil plateau rate decline and water arrival in one of the production wells, PROD3).

The mean of the final $\ln k$ ensemble (Figure 3-6(s)) reproduces the most visible high and low permeability zones, especially around well locations. The mean model is smoother than the reference one as expected: it does not show small scale heterogeneities.

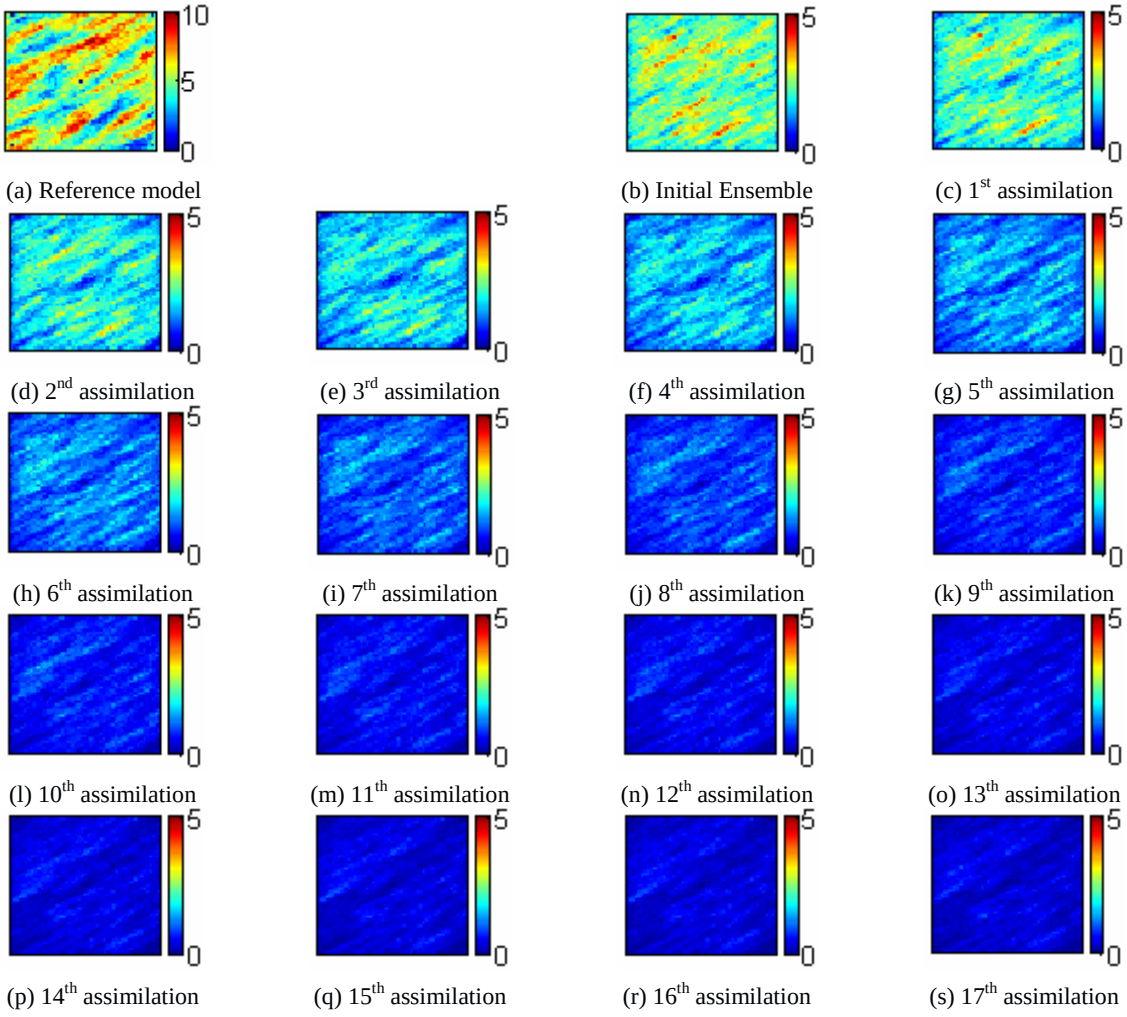


Figure 3-7 (a) Reference Mean model, (b) Variance of the initial ensemble, (c)-(s) Variance of the updated ensemble against assimilation steps.

The variance of the ensemble at each assimilation step is shown in Figure 3-7. When performing assimilations the uncertainty, *i.e.* the model variance, reduces, especially around well locations. As assimilation goes on, variance decreases significantly. This phenomenon is a natural consequence of the assimilation procedure. However, we must avoid too extreme variance reductions. The reason is simple: the ensemble at each time step must be

representative of the possible reservoir models capable of yielding the same trend as the reference model. If the variance is reduced too significantly, all the ensemble members look alike and in the limit, they collapse to a single reservoir model. It is evident that this single final model is only one of the possible reservoir models. This problem can be avoided by increasing the size of the ensemble as shown in section 3.2.1. However, this solution is not always suitable due to practical issues. Improved solutions consist of referring to localizations (section 2.4.1.) or using for instance two parallel ensembles (Houtekamer and Mitchell, 1998).

The convergence of the updated $\ln k$ models toward the reference model can be assessed quantitatively by the Root Mean Square (RMS) error. Two static RMS errors, $RMS_{stat,1}$ and $RMS_{stat,2}$, are considered at each time step. The subscript *stat* stands for static. $RMS_{stat,1}$ considers the difference, grid block by grid block, between each ensemble member and the reference model (equation 3.4). To get $RMS_{stat,2}$, we first compute an average $\ln k$ model over the ensemble at each time step and then consider its departure from the reference model (equation 3.5). These RMS errors are defined as:

$$RMS_{stat,1}(l) = \sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}\frac{1}{N_{xyz}}\sum_{g=1}^{N_{xyz}}(y_{i,g}^{l} - y_{ref,g})^2} \qquad 3.4.$$

and

$$RMS_{stat,2}(l) = \sqrt{\frac{1}{N_{xyz}}\sum_{g=1}^{N_e}(y_{mean,g}^{l} - y_{ref,g})^2} \qquad 3.5.$$

N_e is the ensemble size, N_{xyz} the number of grid blocks, l the assimilation time step, i the ensemble member index and g the grid block index. $y_{i,g}^{l}$ is the value of the static variable (in the case studied, it is $\ln k$) attributed to grid block g for ensemble member i after the l^{th} update. $y_{ref,g}$ and $y_{mean,g}^{l}$ are the reference value and the ensemble mean value updated at time step l for the same grid cell g, respectively. These static RMS errors can be computed for any static variable of interest. These values are then normalized to the initial RMS values. Both of the so defined static RMS errors are represented in Figure 3-8. They decrease as the number of assimilation steps increases meaning that we are getting closer to the reference $\ln k$ model. $RMS_{stat,1}$ decreases more than $RMS_{stat,2}$. Such a result was expected. Indeed, the ensemble mean model at each assimilation step is smoother than each of the ensemble members. Thus, it should differ further from the reference model.

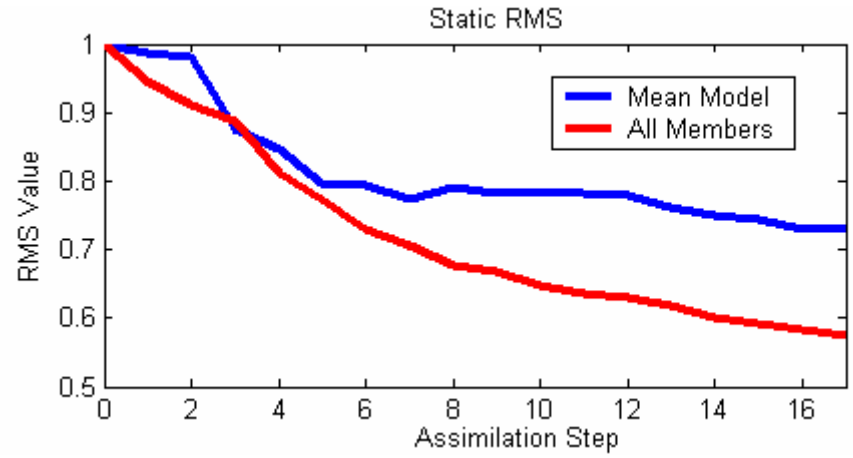


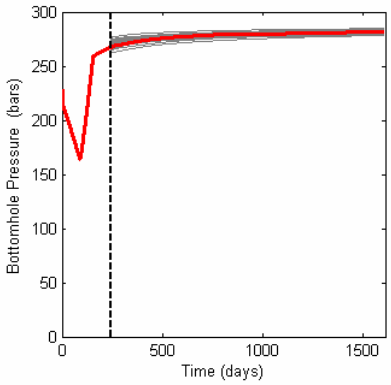
Figure 3-8 Static RMS value computed for ln(k) through assimilations. The red curve corresponds to $RMS_{stat,1}$ and the blue one to $RMS_{stat,2}$.

3.1.4 Production forecasting by the updated models

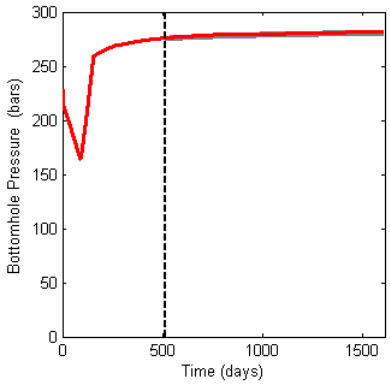
The results presented in Figure 3-3 and Figure 3-4 were obtained by rerunning the whole production history with the permeability models obtained after the 17th assimilation step from time zero. The updated pressure and saturation models were disregarded. In the following discussion, we want to investigate the prediction capability of the full model ($\ln k$, pressure and saturation) after the history-matching phase. Our objectives are two fold. First, we want to determine how many assimilation steps are required to predict accurately well production responses (by accurately, we mean the same production response as the reference model). Second, we want to see whether the predictions should be obtained by rerunning the simulations from time zero or by going on after the last assimilation step. This last alternative would be less expensive in terms of fluid flow simulation.

For that purpose, we consider the predictions obtained with the full updated models after 10 months and 17 months of assimilation.

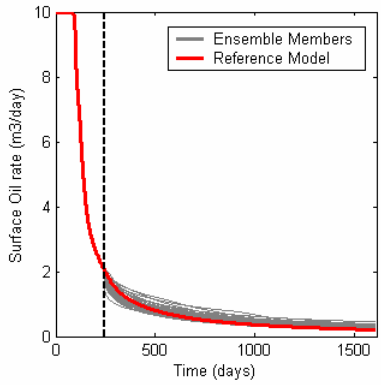
The production responses predicted for well PROD3 are shown in Figure 3-9(a) to (c) for the first case and in Figure 3-9(d) to (f) for the second case. This experiment shows that the predicted models are capable of reproducing the reference well production data for both cases. The only difference is that the spread in the predictions decreases as the number of assimilation steps increases. Consistent with Figure 3-7(l) and (s), the variance of the permeability field after 10 months of assimilation is higher than after 17 months.



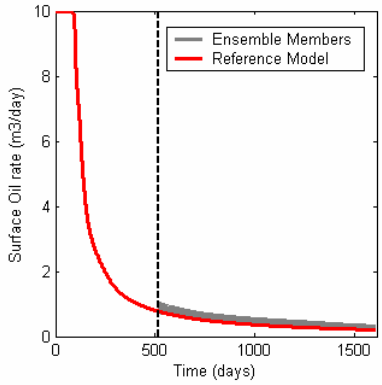
(a) Bottom hole pressure in PROD3 – First Prediction



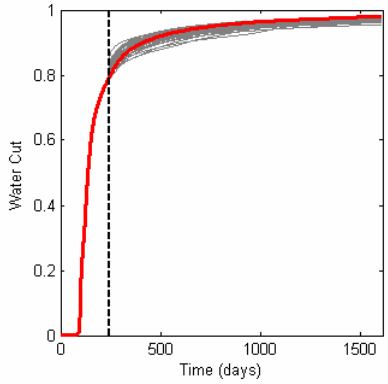
(d) Bottom hole pressure in PROD3 – Second Prediction



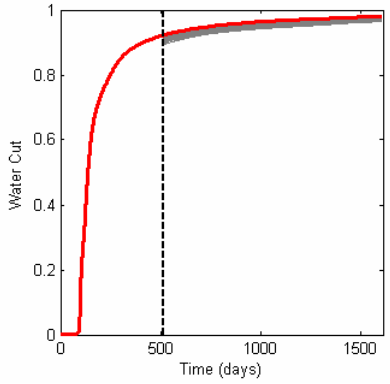
(b) Surface oil rate in PROD – First Prediction



(e) Surface oil rate in PROD3 – Second Prediction



(c) Water cut in PROD3 – First Prediction



(f) Water cut in PROD3 – Second Prediction

Figure 3-9 Prediction of production responses for well PROD3. First column: predictions are computed with updated permeability fields obtained after 10 assimilation steps. Second column: predictions are computed with updated permeability fields obtained after 17 assimilation steps. The red curve represents the reference case. The grey curves correspond to ensemble members. The black dashed line indicates the time limit between the matching period and the forecasting one.

In addition, a closer look at the mean permeability models at these two times (Figure 3-6(l) and Figure 3-6(s), respectively) reveals that most features of the reference permeability field are captured during the first 10 assimilations while only minor additional heterogeneity features are captured afterwards. This explains why similar production behaviors are predicted for the two cases. Hence, it is up to the user to decide how many assimilation steps

should be performed. Although the results are only presented for well PROD3, the same conclusions can be drawn for the other wells.

About the second issue listed above, the quality of the predictions in this experiment is as good as those obtained by rerunning the simulators from time zero (Figure 3-3(d) through (f)). Therefore, in this case, accurate predictions can be obtained by performing cheaper fluid flow simulations after the last assimilation step. These results are of course case dependent.

3.2 Sensitivity studies

The performance of the EnKF can be affected by several controlling parameters such as the ensemble size, the assimilation time step, the level of uncertainty in the measurements and the choice of the initial ensemble. For instance, the low final variance observed above may be linked to the small size of the ensemble. In this section, we perform a set of experiments and aim to assess the performance of the EnKF with respect to individual changes in these parameters. Moreover, we aim to investigate how the EnKF performs when the variogram of the initial ensemble is different from that of the reference model.

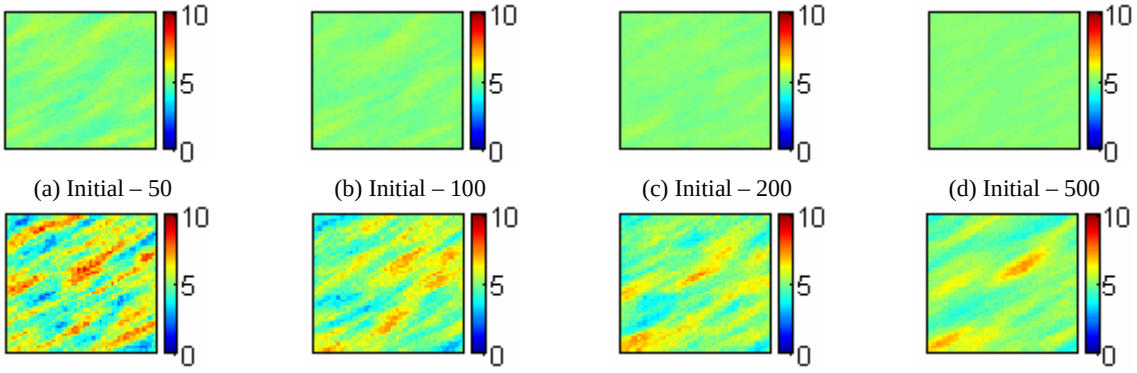
3.2.1 Ensemble size effect

Practically, the EnKF provides an approximation of the error covariance matrix through an ensemble of finite size. As the size N of the ensemble increases, the error on the error covariance matrix decreases by $1/\sqrt{N}$. Therefore, spurious correlations in the covariance matrix can be avoided more efficiently (see section 2.4.1). Clearly, increasing the ensemble size increases the fluid flow simulation and computational overburden. Therefore, there exists a trade-off between the accuracy of the covariance matrix approximation and the overburden. The choice of the ensemble size is case dependent, but in literature an ensemble of size 100 is believed adequate. In this section, we investigate more precisely how many realizations are required in our study to reliably capture the uncertainty in predictions. We apply the EnKF to the same case study as in the previous section, but using ensembles of size 50, 100, 200 and 500. We consider the same data uncertainties and assimilation intervals as above.

Figure 3-10 represents the mean of the initial and final updated ensembles. The mean of the initial models in all cases (Figure 3-10(a) to (d)) is relatively homogeneous. As the size of the ensemble increases, the mean model becomes more homogeneous. The mean of final models

(Figure 3-10(e) to (h)) shows some spatial features similar to the reference model, such as the high permeability regions in the center and low left corner and some of the low permeability zones. As the size of the ensemble increases, the mean of the final models is computed over a larger ensemble and becomes smoother.

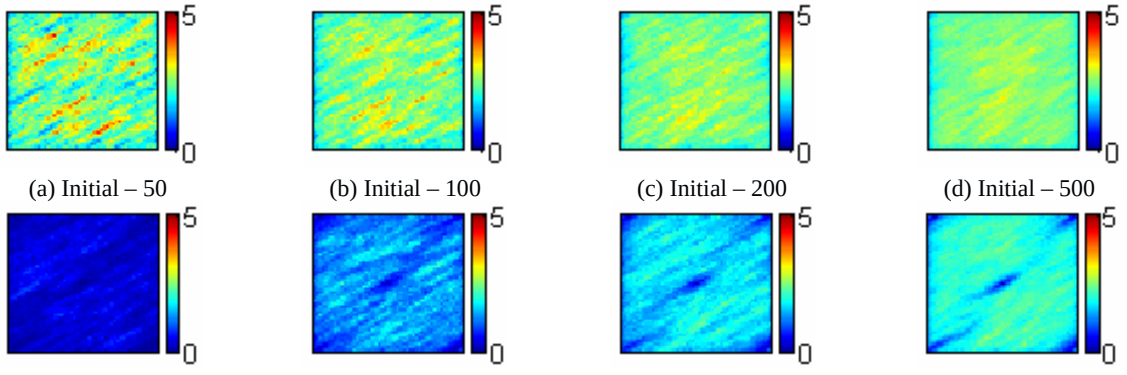
Figure 3-11 represents the variance within each ensemble. The variance of the initial models (Figure 3-11(a) to (d)) is rather homogeneous. Its average value is 2.5, which is the variance for the $\ln k$ model. After data assimilation, the variance of the updated ensembles reduces, especially around well locations. However, its decrease is more severe provided that the ensemble size is small (*e.g.*, 50).



(a) Initial – 50 (b) Initial – 100 (c) Initial – 200 (d) Initial – 500

(e) Final – 50 (f) Final – 100 (g) Final – 200 (h) Final – 500

Figure 3-10 lnk evolution with time. First row: Initial Ensemble Mean for (a) 50 realizations, (b) 100 realizations, (c) 200 realizations, (d) 500 realizations. Second row: Final Ensemble Mean for (e) 50 realizations, (f) 100 realizations, (g) 200 realizations, (h) 500 realizations.



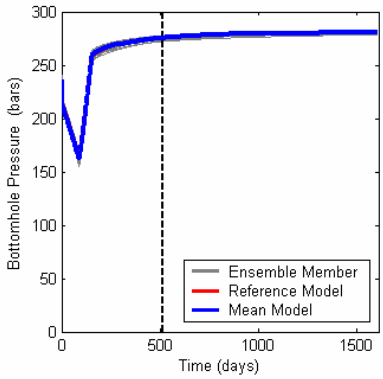
(a) Initial – 50 (b) Initial – 100 (c) Initial – 200 (d) Initial – 500

(e) Final – 50 (f) Final – 100 (g) Final – 200 (h) Final – 500

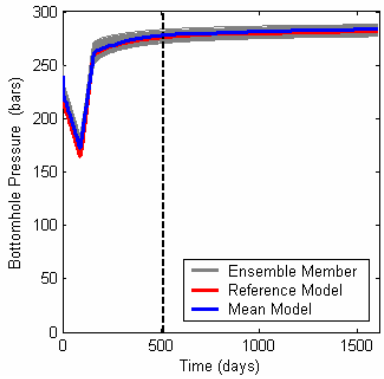
Figure 3-11 lnk evolution with time. First row: Initial Ensemble Variance for (a) 50 realizations, (b) 100 realizations, (c) 200 realizations, (d) 500 realizations. Second row: Final Ensemble Variance for (e) 50 realizations, (f) 100 realizations, (g) 200 realizations, (h) 500 realizations.

We now focus on well production data. The performance of the EnKF for different ensemble sizes at well PROD3 is compared in Figure 3-12, Figure 3-13 and Figure 3-14. Although not presented here, in all cases, the performance of the final ensembles are by far better than those of the initial ensembles. Whatever the ensemble size, the final ensemble members are capable

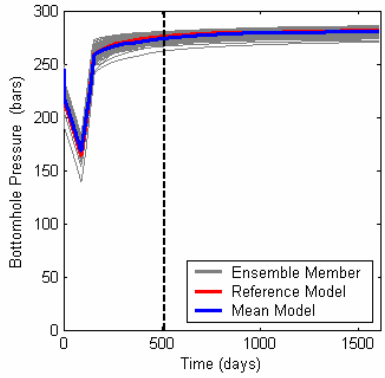
of following the trend of the reference production data (represented by the red line) within the 5% relative measurement error, in both the history-matching and the forecasting periods. Therefore, the matching process is considered as successful.



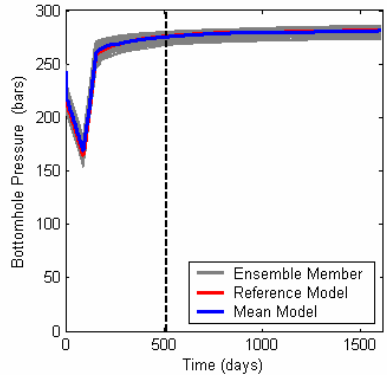
(a) Bottom hole pressure in PROD3 – 50 Members



(b) Bottom hole pressure in PROD3 – 100 Members



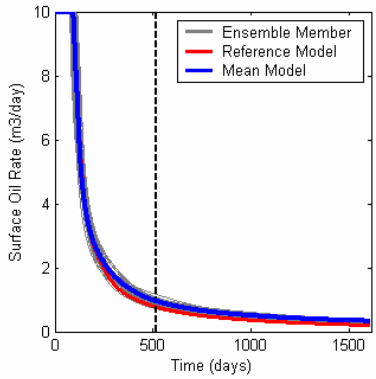
(c) Bottom hole pressure in PROD3 – 200 Members



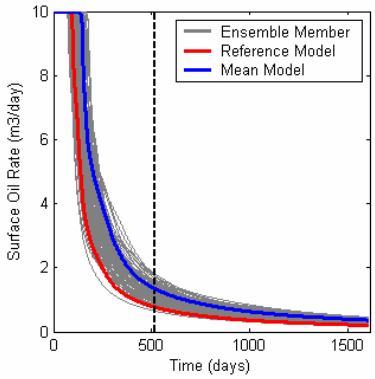
(d) Bottom hole pressure in PROD3 – 500 Members

Figure 3-12 Bottom hole pressure at Well PROD3 for Final Ensembles of size (a) 50, (b) 100, (c) 200, (d) 500. The black dashed line separates the matching (0-516 days) and forecasting (517-1612 days) periods.

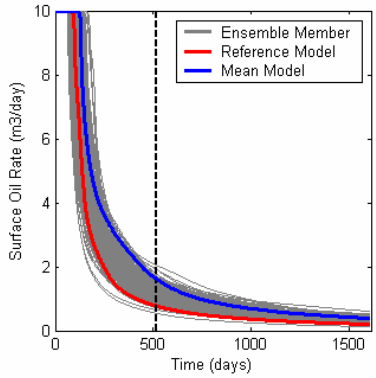
In addition, when increasing the size of the ensemble, the spread within each of the production responses increases after the history-matching period. This fact is a direct consequence of the variance preservation property in larger sized ensembles. This effect can be better observed for the bottom hole pressure (Figure 3-12). The same effect is observed for the surface oil rate and water cut at the beginning of the forecasting period, but progressing toward the end of this period, this effect diminishes as these two values reach their lowest and highest limits, respectively. These results are more pronounced when the size of the ensembles increases from 50 to 100. The improvements by using larger ensemble sizes, *i.e.* 200 and 500, are less significant. Therefore, an ensemble of size 100 seems appropriate for this study. However, still, more variance is observed for larger ensemble sizes.



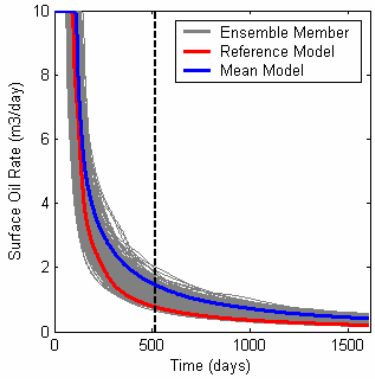
(a) Surface oil rate in PROD3 – 50 Members



(b) Surface oil rate in PROD3 – 100 Members

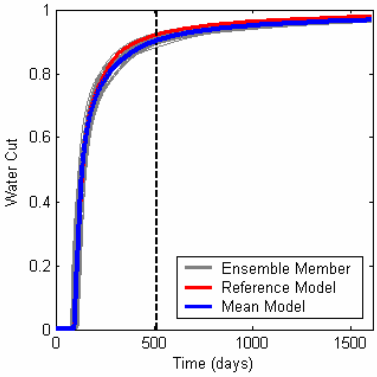


(c) Surface oil rate in PROD3 – 200 Members

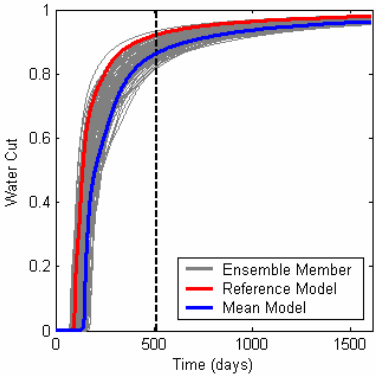


(d) Surface oil rate in PROD3 – 500 Members

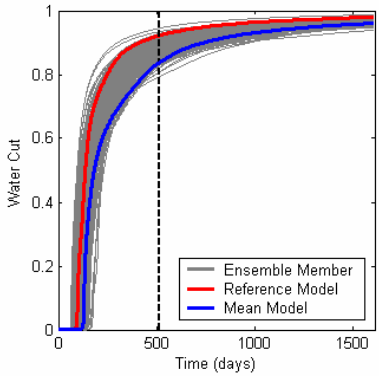
Figure 3-13 Surface oil rate at Well PROD3 for Final Ensembles of size (a) 50, (b) 100, (c) 200, (d) 500. The black dashed line separates the matching (0-516 days) and forecasting (517-1612 days) periods.



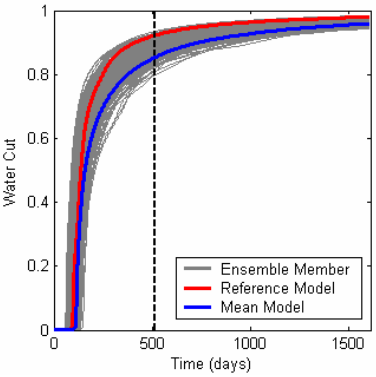
(a) Water cut in PROD3 – 50 Members



(b) Water cut in PROD3 – 100 Members



(c) Water cut in PROD3 – 200 Members



(d) Water cut in PROD3 – 500 Members

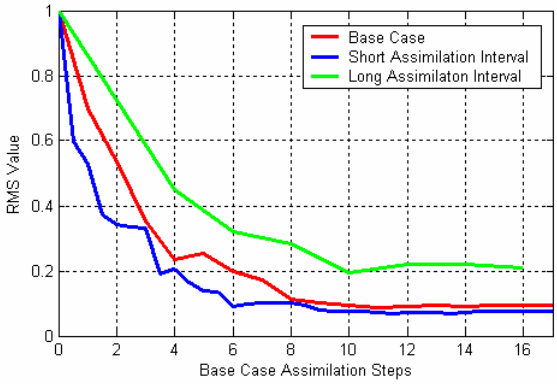
Figure 3-14 Water cut at Well PROD3 for Final Ensembles of size (a) 50, (b) 100, (c) 200, (d) 500. The black dashed line separates the matching (0-516 days) and forecasting (517-1612 days) periods.

3.2.2 Assimilation time step effect

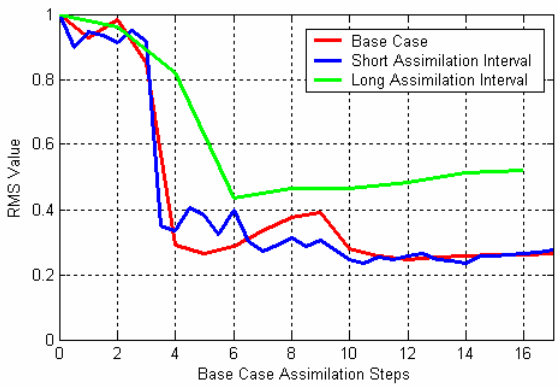
The time interval between two consecutive EnKF assimilations is an important issue for practical applications of the EnKF. In general, data assimilation is required whenever significant flow behavior changes occur in the reservoir such as opening or shutting existing wells, adding new wells or even changing the production and injection rates. The assimilation time interval should not be too big or too short. In the latter case, the production data could be correlated and redundant, thus no more information can be withdrawn. However, non linearity in the flow behavior may not be captured for too time-spaced assimilations.

In order to investigate the effect of the assimilation time interval on the quality of the updated models by the EnKF, we investigate two scenarios and compare the results with the "base case" presented in section 3.1.2. In the first scenario, termed as "short assimilation interval", two assimilations are performed each month, *i.e.*, the middle and the last day of each month. In the second scenario termed as "long assimilation interval", one assimilation is performed every two months. For the short assimilation interval run, there are 34 assimilations in total, whereas for the long assimilation interval run, there are only 8 assimilations over 16 months. For the base case, one assimilation is performed each month.

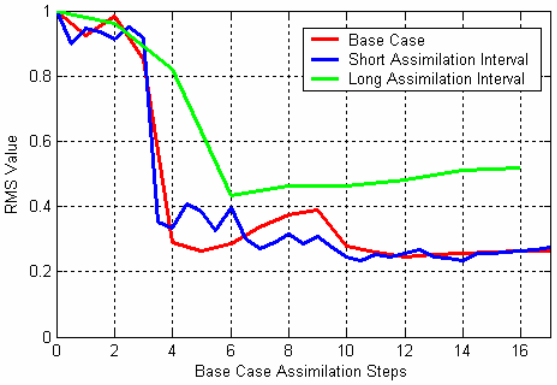
The $RMS_{prod,2}$ on the mean model (equation 3.3) is displayed in Figure 3-15 for each of the well production data. It is computed for the three pre-defined assimilation intervals. For the bottom hole pressure (Figure 3-15(a)) the best RMS results are those of the short assimilation interval, followed by the base case and then the long assimilation interval, as expected. The same trend, in general, is observed for surface oil rate (Figure 3-15(b)) and water cut (Figure 3-15(c)). However, a closer look reveals other features related to the assimilation interval. The RMS for the surface oil rate or water cut shows an increasing trend for the long assimilation interval after the 6^{th} assimilation step. At this time, the injection rate undergoes its second incremental increase, which strongly impacts fluid flow. The same type of behavior is observed for the base case. However, the increasing trend dampens after a certain number of assimilations. Interestingly, this anomaly disappears in the case of short assimilation interval, indicating that problems associated to nonlinearities in fluid flow can be compensated for by decreasing the time intervals.



(a) RMS on Bottom hole pressure



(b) RMS on Surface oil rate



(c) RMS on Water cut

Figure 3-15 Dynamic $RMS_{prod,2}$ relative to (a) Bottom hole pressure, (b) Surface oil rate, (c) Water cut against the number of assimilation for the three assimilation interval scenarios.

3.2.3 Measurement uncertainty effect

Measurement uncertainty, *i.e.* standard deviation of the measurement noise or error, in a history-matching problem depends on the uncertainty in the measurement devices, their positioning, *etc.* Thus, it is case-dependent and should be fixed at the beginning of the assimilation process. Meanwhile, the same uncertainty levels are used to generate the ensemble of perturbed measurements. This case study is a synthetic one and it allows us to test the sensitivity of the EnKF method to measurement uncertainties.

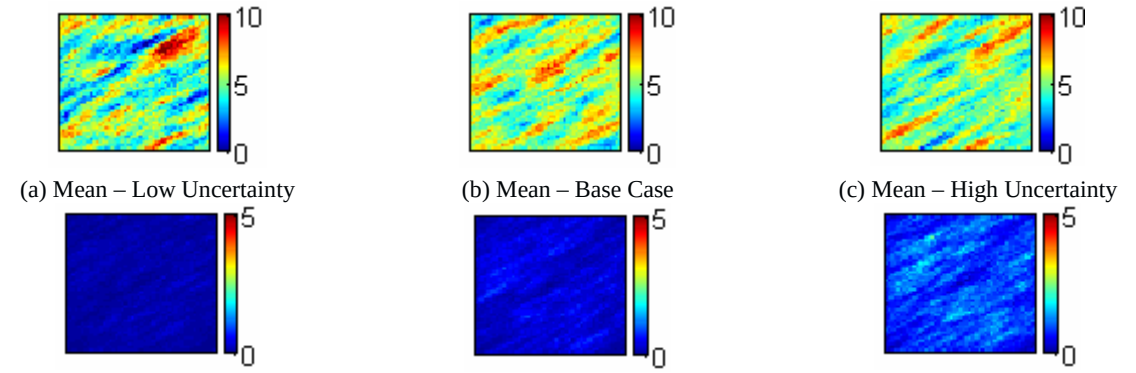
In an independent set of runs, the ensemble used in section 3.1 is assimilated with three levels of measurement uncertainty: low, base and high.

Table 3-2 The Standard deviation of error for well production data relative to the real measurements

Standard Deviation	Low (SD,%)	Base (SD,%)	High (SD,%)
BHP	0.01	0.05	0.1
SOR	0.01	0.05	0.1
WCT	0.01	0.05	0.1

These are summarized in Table 3-2. The assimilation scenario is the one described in section 3.1.1 for the three cases.

In this experiment, we first aim to have a look at the final matched $\ln k$ models. In all the three cases, the initial ensemble of section 3.1 is utilized with the initial mean and uncertainty given in Figure 3-6(b) and Figure 3-7(b). As can be seen in Figure 3-16, all models regenerate after assimilations the main features of the reference case, but with smoother results as uncertainty increases. Although we started with the same level of variability within ensemble members for all the three cases, the lower the measurement uncertainty, the higher the variance reduction. The low uncertainty case is very close to the one where measurements are not treated as random functions, *i.e.* if we do not consider an ensemble of measurements. Our results confirm that the final variability of the ensemble is too low in that case (Burger *et al.*, 1998).

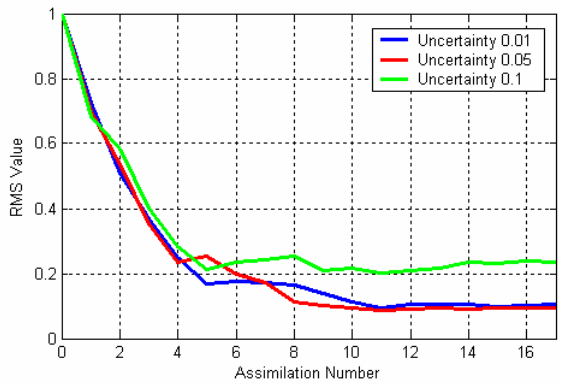


(a) Mean – Low Uncertainty (b) Mean – Base Case (c) Mean – High Uncertainty

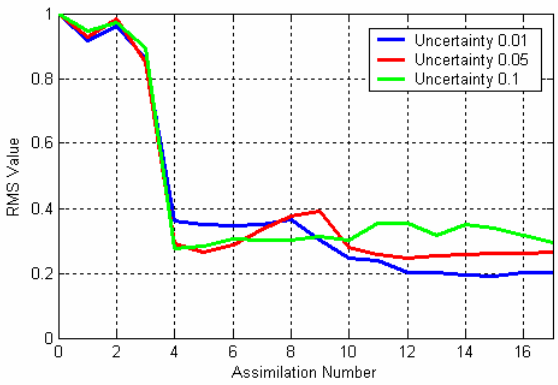
(d) Variance – Low Uncertainty (e) Variance – Base Case (f) Variance – High Uncertainty

Figure 3-16 The lnk for the final assimilation. First row: final Ensemble Mean for (a) low Uncertainty, (b) base Case, (c) high Uncertainty. Second row: final Ensemble Variance for (d) low Uncertainty, (e) base Case, (f) high Uncertainty

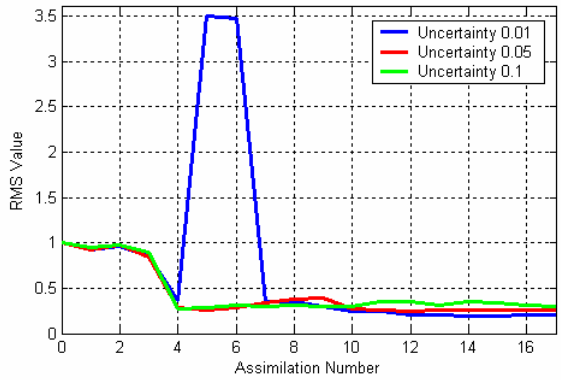
The dynamic $RMS_{prod,2}$ for the mean model is presented for each of the production data in Figure 3-17. A decreasing trend is observed for all cases. The most significant decrease is obtained for the low uncertainty case, followed by the base case and the high uncertainty case. The differences are not significant, but the lowest uncertainty case exhibits some anomalies for the water cut RMS, which damps as assimilation continues. This anomaly is observed at the time step corresponding to the assimilation of the first water cut value. It seems that the reduced variance within the ensemble of the low uncertainty cannot regenerate this phenomenon at first try. However, later, with more assimilation, the reference water cut is regenerated.



(a) RMS for Bottom hole pressure



(b) RMS for Surface oil rate



(c) RMS for Water cut

Figure 3-17 Dynamic $RMS_{prod,2}$ for (a) Bottom hole pressure, (b) Surface oil rate, (c) Water cut against the number of assimilations with various measurement uncertainty levels.

Although the appropriate choice for measurement uncertainty depends on measurement devices, a too low uncertainty may pose problems in the workflow of the EnKF, as it can induce an excessive variance reduction. If this problem is avoided, our experiment suggests that the choice of measurement uncertainty is of less importance. Meanwhile, the base and high uncertainty cases provide improved estimates of permeability fields and match the production data fairly well.

3.2.4 Spatial variability preservation

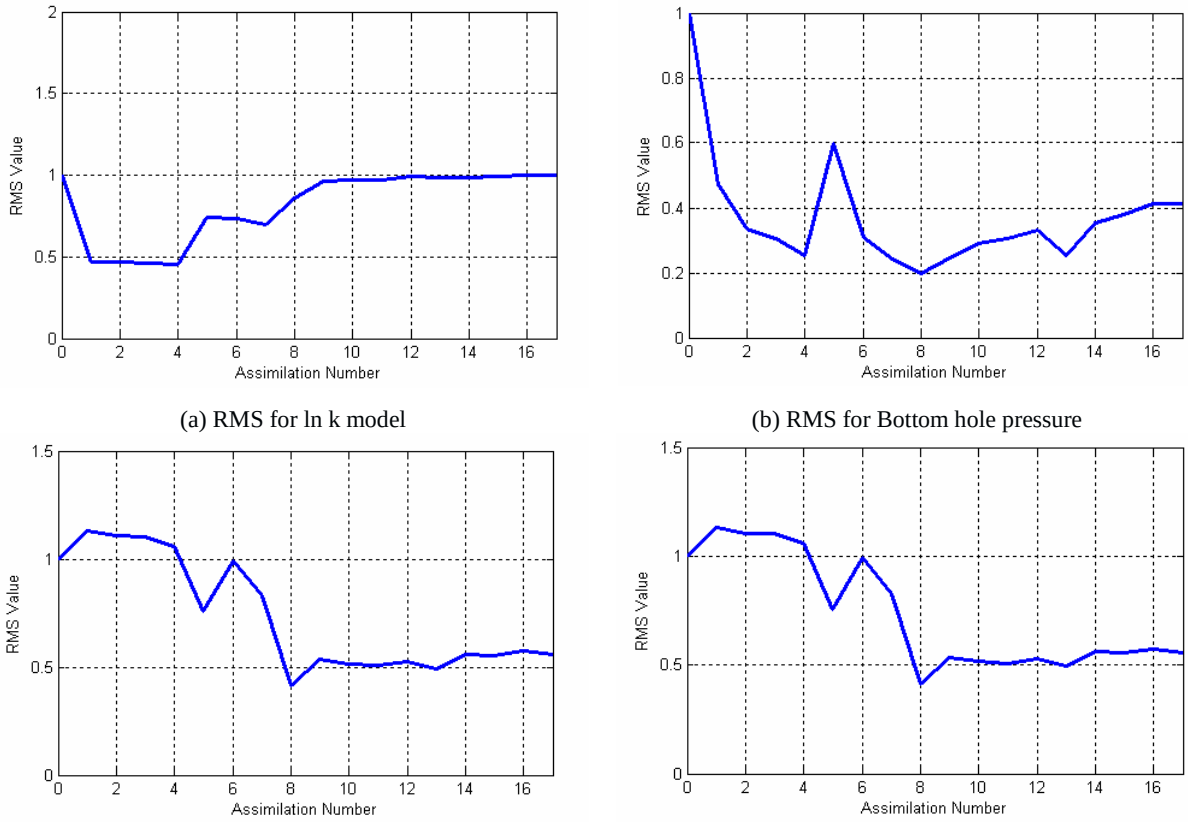
We are interested in investigating the potential of the EnKF to capture the spatial variability of the reference model when starting from an ensemble whose members are characterized by a spatial variability different from the one of the reference model. For simplicity, the spatial variability of $\ln k$ is assumed to be characterized by the mean and covariance. Thus, we keep the same reference model as above, but generate the initial ensemble with the following spatial properties:

Mean: 5.0, Variance: 2.5, Variogram: Spherical (Isotropic), Correlation length: 200m.

(a) Reference Model (b) Mean of the Final Ensemble

Figure 3-18 ln k model (a) Reference Model, (b) Final updated mean model with a new initial ensemble.

The EnKF method is run to get the final updated models. Their mean is shown in Figure 3-18. It looks quite different from the reference model. Moreover, the anisotropy of the reference was not retrieved and the spatial variability of the mean final models remains close to the one of the initial ensemble.



(a) RMS for ln k model (b) RMS for Bottom hole pressure

(c) RMS for Surface oil rate (d) RMS for Water cut

Figure 3-19 RMS against the number of assimilation for (a) ln k Mean Model, (b) Bottom hole pressure, (c) Surface oil rate, (d) Water Cut. The initial ensemble consists of ln k models with an isotropic spherical variogram.

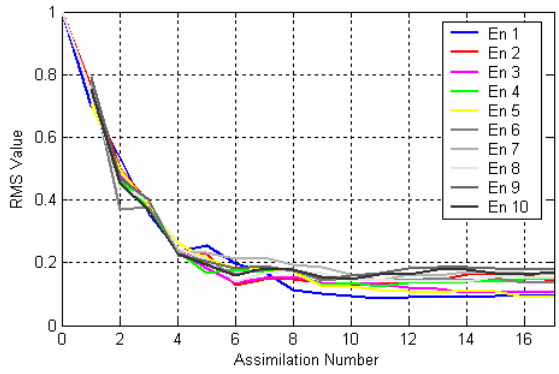
Next, we compare both $RMS_{prod,2}$ and $RMS_{stat,2}$ (equation 3.3 and 3.5, respectively) for the three studied cases. Figure 3-19(a) shows that as assimilations go on, the mean ensemble model first converges and then deviates from the reference model. The final model is not better than the initial one. In spite of the poor static RMS results, the $RMS_{prod,2}$ shows a general decreasing trend, but the results are not as satisfactory as for the base case. This

indicates that although there may exist different reservoir models that match the production data fairly well, they may fail to regenerate the reference model.

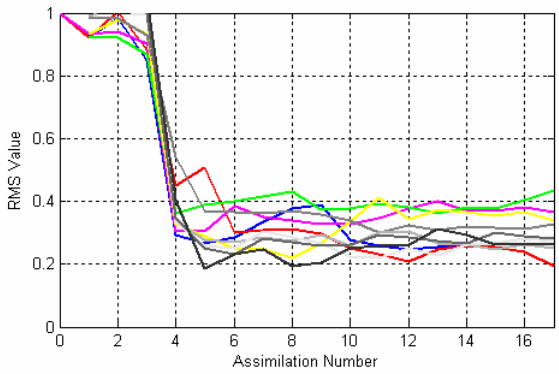
However, it is expected that when more data are available, the additional assimilation steps contribute to better capture the spatial variability of the reference permeability field.

3.2.5 Initial ensemble uncertainty effect

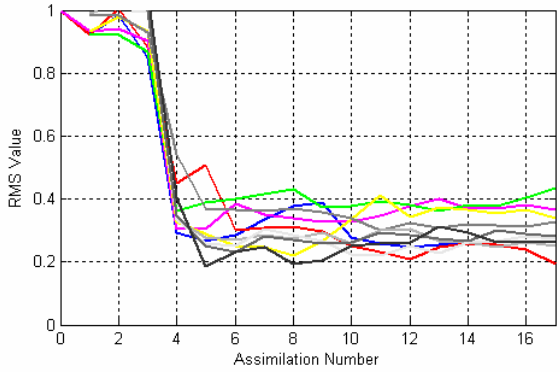
One of the main issues with the EnKF is the choice of an initial ensemble which represents accurately the space of potential initial reservoir models. In addition, due to lack of independence between ensemble members, one cannot calculate the uncertainty in the estimated posterior distribution, given only a single EnKF run (Thulin *et al.*, 2008). The idea in this section is to investigate and compare how different ensembles of the same size can affect the performance of the EnKF. The study is mainly based on the prediction of cumulative oil production during the forecast period. To this end, ten different ensembles of size 50 are used to perform the assimilation procedure described in section 3.1.



(a) RMS on Bottom hole pressure



(b) RMS on Surface oil rate

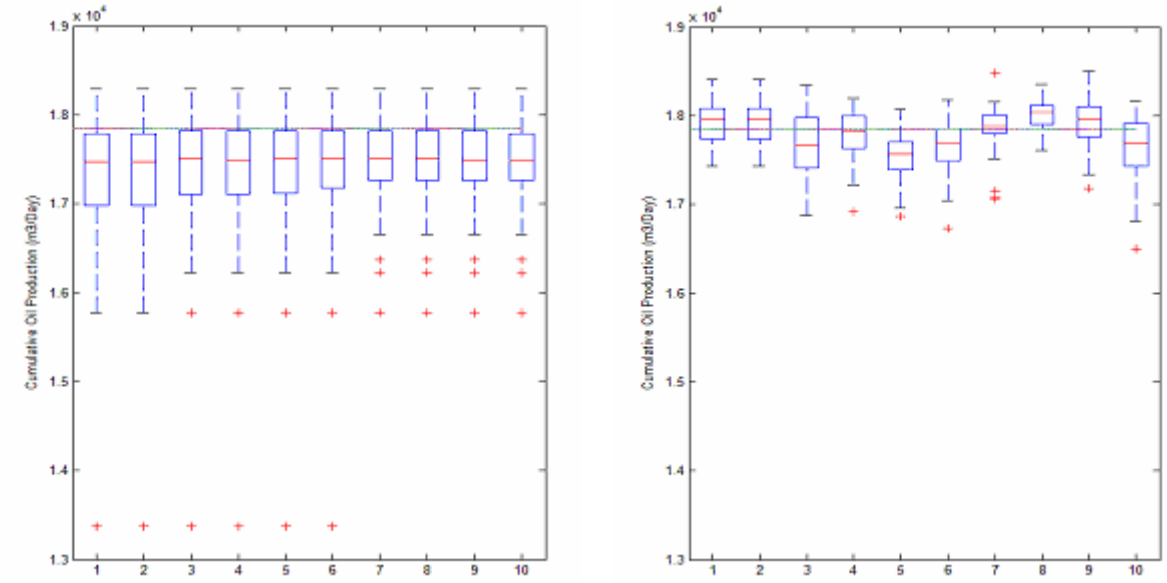


(c) RMS on Water cut

Figure 3-20 Dynamic $RMS_{prod,2}$ profile of (a) Bottom hole pressure, (b) Surface oil rate, (c) Water cut throughout the Assimilations for the Set of Initial Ensembles.

Figure 3-20(a) to (c) represent the dynamic $RMS_{prod,2}$ results for each of the well production data starting from 10 different initial ensembles. In all cases, the RMS decreases. For the bottom hole pressure, the RMS results are very close, but for the surface oil rate and water cut, more differences can be observed. Moreover, for these two production responses, several anomalies, *i.e.* slight increasing trends, are present which damp down as assimilations continue. This artifact could be avoided by decreasing the assimilation time interval as presented in section 3.2.2.

Then, we compare the forecast capability of each of the ensembles. To do so, the total cumulative oil production predicted at the end of the forecast period from the initial and final models (obtained after 17 months of assimilation) are computed and represented by the box plots in Figure 3-21.



(a) Initial Ensembles (b) Final Ensembles

Figure 3-21 Box plot of Cumulative Oil Production for the 10 Ensembles (a) Initial Ensembles (b) Final Ensembles

In this figure, the horizontal line shows the cumulative oil production estimated for the reference model after 1612 days of production. All initial ensembles embrace the reference model within their extent beyond the upper quartile, but have outliers with a cumulative oil production much lower than the reference. After assimilations the variation in the cumulative oil production between ensemble members is reduced (Figure 3-21(b)). For most of the ensembles, except ensembles 5 and 8, the true value is between the lower and the upper quartiles. Furthermore, the outliers are removed in ensembles 1, 2, 3 and 8. The ones that still exist are not far from the lower extent of the data. For all ensemble members, the difference between the median of the ensemble and the true value is reduced, indicating that the mean is

an acceptable estimator within the ensemble. Our results suggest that the range of predictions is more reliable whenever the test is repeated for several ensembles of the same size.

3.3 Conclusions

In this chapter, we investigated the performance of the EnKF as a history-matching method for a 2D synthetic reservoir model. Comparing initial and final updated ensembles, after several assimilations, the spread of the well production responses was reduced around the reference model and all the ensemble members were capable of tracking its trend. The evolution of the $\ln k$ model over time was represented: most high and low permeability features of the reference model were retrieved by the end of the assimilation period. The results were also assessed qualitatively by static and dynamic RMS values (Section 3.1). It was observed that non linearity in flow behavior may induce anomalies in the decreasing trend of the RMS during assimilations.

Section 3.2 presents a set of sensitivity tests to determine the effect of several controlling parameters. First, the effect of ensemble size was studied. By increasing the ensemble size, little is gained in the quality of the production data match, but the variation among ensemble members is preserved throughout assimilation and the flow responses simulated for the final updated models show more variation (Section 3.2.1). Second, the impact of the assimilation interval was investigated. The smaller the assimilation interval, the better the match in production data (Section 3.2.2). Third, three levels of measurement uncertainty were considered in the assimilation procedure. The results indicate that a too low uncertainty is unfavorable as the variance reduction is great. However, as a general rule the sensitivity of the EnKF to this parameter is fairly less prominent than the others (Section 3.2.3). Forth, we studied the dependence of the EnKF to the initial ensemble in terms of geostatistical parameters, in our case, anisotropy. It was observed that although the production data were matched adequately, this experience was not capable of regenerating the reference model (Section 3.2.4). Last, Section 3.2.5 dealt with the capability of different ensembles of the same size in predicting the total cumulative oil production through a forecasting period. It was concluded that several ensembles of the same size are required to make a sound forecast with ensembles of fairly small size. In petroleum applications of the EnKF, the notion of model error is disregarded. However, its inclusion may be considered as a solution for significant variance reduction within ensembles of small size.

4 Variant of PUNQ-S3 Case Study

The 2D simple synthetic case considered in the previous chapter enables us to assess the performance of the EnKF and investigate the influence of parameters involved in its formulation. However, to process the merits and shortcomings of this method, its performance should be verified for several synthetic and real case studies. In this chapter, we consider a more complex case study for assimilation derived from the PUNQ-S3 case. The reservoir model and the EnKF setting considered to perform the history-matching of well production data are described in Section 4.1. Assimilation results are given in Section 4.2. Section 4.3 is devoted to sensitivity tests to assess the influence of several parameters, *i.e.* ensemble size, assimilation time interval and measurement uncertainty level. The experiments in this chapter provide the EnKF user with a more robust insight on this methodology. More details can be found in Heidari *et al.* (2010).

4.1 Overview of the PUNQ-S3 case

The PUNQ-S3 case study is a standard small-size reservoir engineering model set up by the PUNQ project (Production forecasting with UNcertainty Quantification) and often used for performing benchmarks. It is based upon a real field that has been operated by Elf Exploration and Production. A full description of this case study can be found on the PUNQ webpage and in Floris *et al*. (2001).

4.1.1 Geological description

The PUNQ model encompasses 5 layers with different petrophysical properties due to various depositional environments, whose main characteristics are summarized as follows:

- Layers 1, 3 and 5 correspond to fluvial channels encased in floodplain mudstone. They consist of a low porosity shale matrix (porosity < 5%) with linear streaks of high porosity sand (porosity > 20%). These two sandy and shaly facies are represented by an "effective" facies with good reservoir properties.

- Layer 2 consists of marine or lagoonal shales with distal mouthbars. This results in a low porosity shaly matrix (porosity < 5 %) with a few higher porosity patches. Again, an "effective" facies with poor reservoir properties is used to represent this layer.
- Layer 4 is a lagoonal delta, encased in lagoonal clay. It results in a low porosity matrix (porosity < 5%) with a more intermediate porosity region (porosity ~ 15%). This layer is then populated by an "effective" facies with intermediate reservoir properties.

4.1.2 Reservoir model

The numerical model is built over a 19×28×5 grid. The dimensions of each grid block in the X and Y directions are 180×180 m^2 and the thickness is defined according to the database data set. There are 1761 active cells and the reservoir is produced from 6 wells named PRO-1, PRO-4, PRO-5, PRO-11, PRO-12 and PRO-15. Figure 4-1 displays the top structure (layer 5) and the 6 well locations.

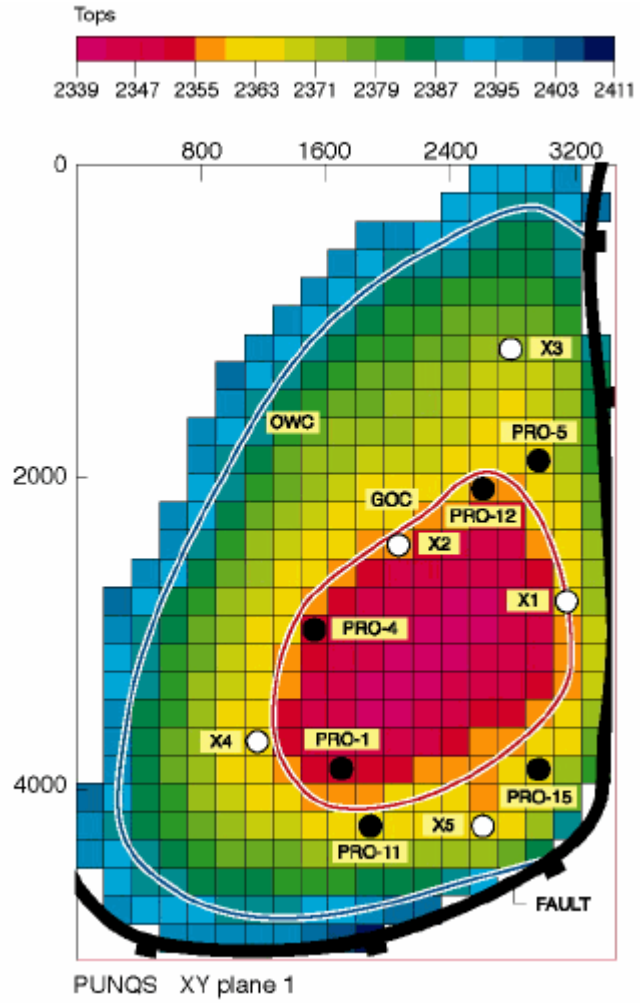


Figure 4-1 Top structure and well locations – PUNQ-S3 model (from PUNQ webpage)

The field is bounded by a fault to the east and south and by a strong aquifer to the west and north. Due to the strength of the aquifer, injection wells were not needed for pressure maintenance.

4.1.3 Petrophysical properties

Porosity and horizontal permeability realizations are stochastically drawn to populate the five layers of the reservoir and their statistical properties are reported in Table 4-1. Vertical permeability is assumed to be the same as the horizontal one. The data used to generate the realizations differed slightly from those provided by the PUNQ webpage, in order to enlarge the possible variations when performing sensitivity tests with the EnKF. In addition, the spatial variations in porosity and permeability are characterized by a spherical variogram, whose main axes and anisotropy ratios are given in Table 4-1.

Table 4-1 : Facies properties – Porosity, logarithm of horizontal permeability ln (kh) and variogram data.

Layer	**1**	**2**	**3**	**4**	**5**
Facies	**A**	**B**	**C**	**D**	**E**
Porosity Mean	0.1722	0.0802	0.1677	0.1615	0.1892
Porosity Variance	0.0078	0.0004	0.0050	0.0006	0.0049
ln (kh) Mean	2.18	1.41	2.24	2.47	2.49
ln (kh) Variance	3.14	0.74	3.26	5.64	3.72
Correlation Length (m)	3500	750	6000	1500	3750
Facies Anisotropy Ratio	0.286	1.0	0.25	0.50	0.333
Azimuth (degrees)	30	0	45	-30	60

All realizations (including the reference reservoir model) were generated on the basis of the Fast-Fourier Transform-Moving average (FFT-MA) algorithm (Le Ravalec *et al.*, 2000) without any conditioning to data at well locations. The resulting porosity and permeability models, which are used as the reference model in this study are displayed in Figure 4-2 and Figure 4-3.

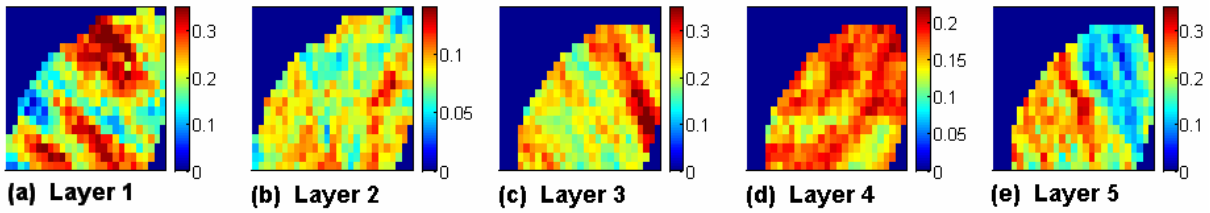


Figure 4-2 Reference porosity (in fraction) fields for the 5 layers in the model.

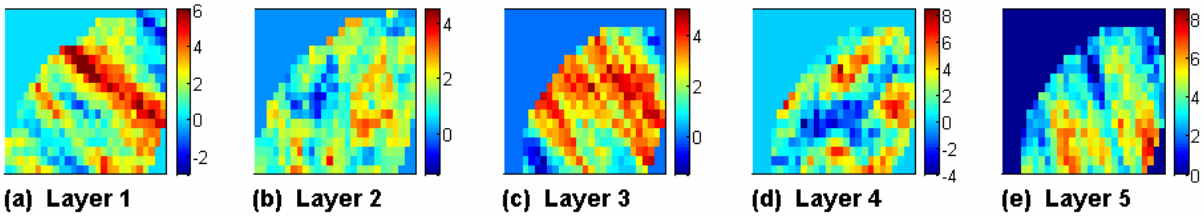


Figure 4-3 Reference permeability, ln(kh), fields for the 5 layers in the model.

4.1.4 Production schedule

The original PUNQ-S3 case involves 3-phase flows in the reservoir. In the present work, we consider a simpler 2-phase (oil and water) black-oil case, which has the advantage to provide more easily readable results. The original PVT data were modified to account for this change, while aquifer data remained unchanged. Relative permeability was generated according to the charts available in our database and capillary pressure was assumed to be negligible. We kept a production schedule similar to the one developed for the original model, which is the same regardless of the production well. It consists of the following phases (Figure 4-4):

- 1 year of extended well testing with:
  - four three-monthly production periods with oil production rates of 100, 200, 100, 50 m^3/day, respectively;
- 3 years of well shut-in;
- 4 years of production.
  - A well shut-in test is performed during the last two weeks every year to collect shut-in pressure data. During the rest of the year, a constant production rate of 100 m^3/day is set up. This flow rate constraint switches to a bottom-hole pressure (BHP) constraint when pressure drops down to 120 bars.

The study time period spans between 01/01/1967 and 15/01/1975, *i.e.* 8 years and it is referred as the *history-matching interval* in the remainder of this chapter. The time period is extended to 01/07/1983 to perform 8.5 years of forecasting and this additional period is termed the *prediction interval*. During the history-matching interval, measurements are obtained at irregular intervals. The noise level, *i.e.*, the standard deviations of the normal distribution defining the measurement uncertainties, were set to 2 bars for pressure, 2 m^3/day for surface oil rate and 0.02 for water cut.

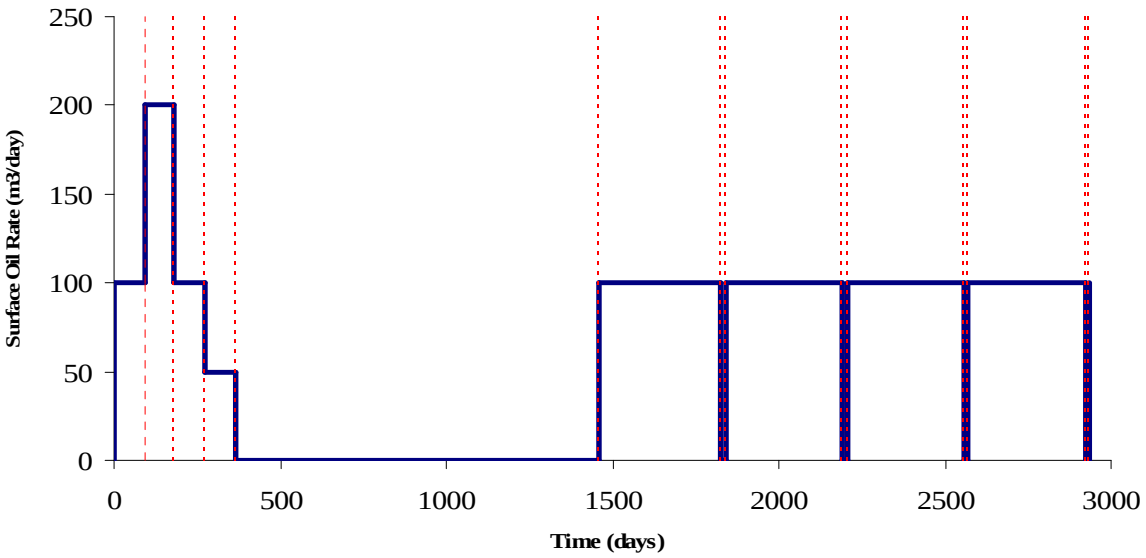


Figure 4-4 Evolution of the surface oil rate scheduled at the production wells during the history-matching interval. The vertical red lines indicate the assimilation times.

4.2 History-matching of the reference case

We designed a twin experiment to investigate the usefulness of the EnKF for performing history-matching on our modified version of the PUNQ-S3 case. We followed the production schedule described above to simulate production data during the history-matching interval for our reference reservoir model (Figure 4-2 and Figure 4-3): bottom hole pressure (BHP), surface oil rate (SOR) and water cut (WCT). The resulting values obtained for the 6 target producers yield the observations that need to be reproduced through the history-matching process. They will be referred to a "reference production data" in the following. Values obtained for wells PRO-5 and PRO-15 are shown in Figure 4-5 and Figure 4-6 respectively (red curves).

The reference porosity and permeability fields are now assumed to be unknown contrary to the reference production data. The primary purpose of the EnKF approach is then to identify a set of realizations conditioned to these production data.

The history-matching interval is split into 13 successive time steps, the end of each time step being associated to a change in the flow rate. These intervals or assimilation steps are displayed in Figure 4-4 on top of the expected surface oil rate variations against time (red lines). For each time step, there are 18 reference data available: 6 BHPs, 6 SORs and 6 WCTs.

The size of the ensemble is set to 50: for each of the layers in the model, 50 realizations are generated for both porosity and permeability. These realizations have the same statistical properties (*e.g.* mean and variogram) as the reference reservoir model, while the other parameters such as the relative permeability curves, the initial pressures and the initial water saturations are assumed to be known without uncertainty. Each ensemble member is updated according to the assimilation timing defined above. This ensemble is denoted *Ensemble 1* in the following.

As the number of active cells is 1761, there are 1761 unknown porosity and permeability values. Similarly, there are also 1761 pressure and water saturation values. Permeability and porosity are deemed as static parameters while water saturation and pressure are dynamic variables, so that the state vector for the j^{th} ensemble member at the k^{th} time step is defined by:

$$\psi_{k,j} = [trans(\varphi_{k,j,1}),...,trans(\varphi_{k,j,N_a}),trans(kh_{k,j,1}),...,trans(kh_{k,j,N_a}), \quad 4.1.$$
$$P_{k,j,1},...,P_{k,j,N_a},S_{wk,j,1},...,S_{wk,j,N_a},d_{k,j,1},...,d_{k,j,N_d}]$$

N_a is the total number of active grid blocks in the reservoir model and N_d the total number of data. $\varphi_{k,j,i}$, $kh_{k,j,i}$, $P_{k,j,i}$ and $S_{wk,j,i}$ are the porosity, horizontal permeability, pressure and water saturation in grid block i, respectively. $d_{k,j,i}$ stands for the i[th] reference production data, *i.e.*, BHP, SOR or WCT for the 6 production wells. Function *trans* defines a transformation applied to state parameters, which ensures that these parameters still have physical values at the end of the updating step. These transformations are:

1. for porosity $\quad trans(\varphi) = \log\left(\frac{\varphi}{1-\varphi}\right)$
2. for permeability $\quad trans(kh) = \log(kh)$

4.2.1 Results for production data

We now compare the performance of the initial and corrected models with regards to the simulated response, *i.e.* BHP, SOR and WCT, and focus on two wells: well PRO-5 and well PRO-15 (Figure 4-5 and Figure 4-6). The production responses are simulated for the history-matching interval (0-2936 days) and the prediction interval (2936-6025 days). For the final ensemble, each updated model is taken back to day 0 and numerically simulated for the whole time span.

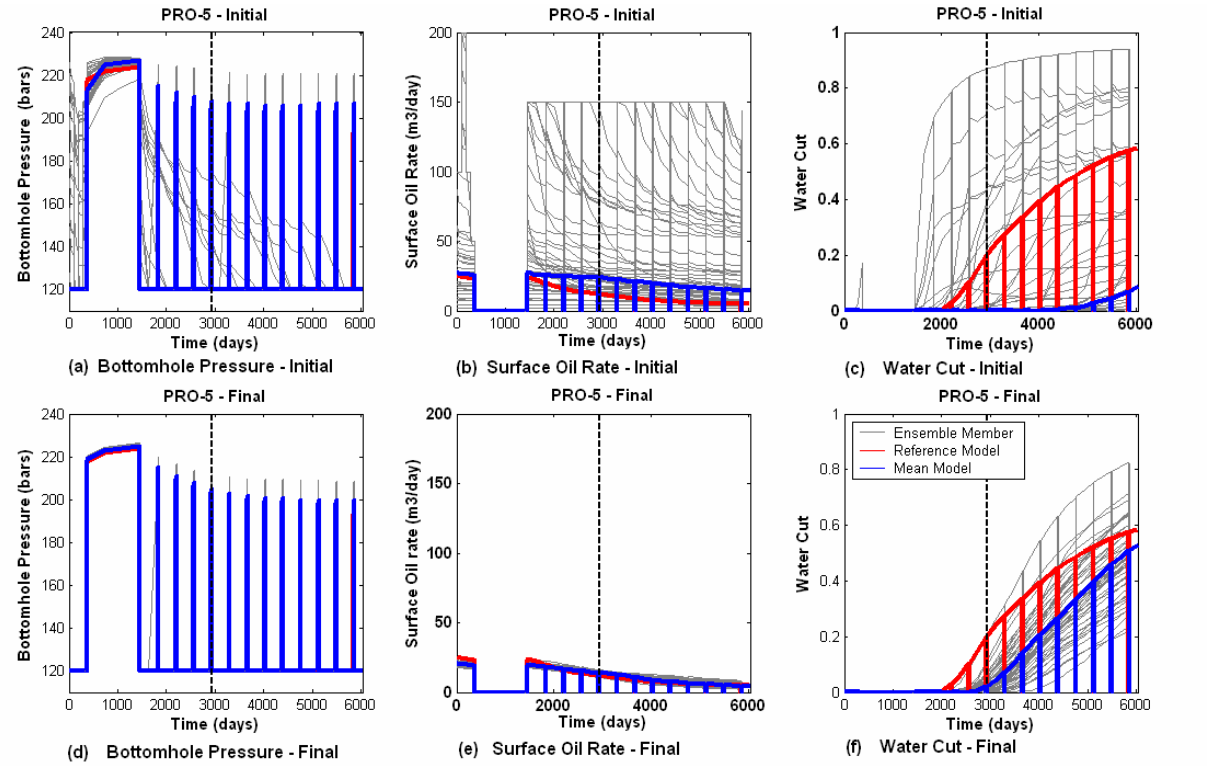


Figure 4-5 Production data simulated for well PRO-5 with an ensemble of size 50 during the history-matching and prediction periods. First row: Initial ensemble. Second row: Final ensemble updated with EnKF. The black dashed vertical line at 2936 days indicates the time limit between the history-matching interval (0-2936 days) and the prediction interval (2936-6025 days). The gray curves correspond to the ensemble members, the red curves to the reference model and the blue curves to the results obtained with the reservoir model computed as the mean of the ensemble.

Comparison of the two rows of Figure 4-5 and Figure 4-6, respectively, stresses that the production responses simulated for the initial models fluctuate much more than those obtained for the corrected models. Moreover, the initial models do not follow the trend of the reference model and the mean reservoir model derived from this initial ensemble behaves very differently from the reference model. However, after correction with the EnKF, the difference between the ensemble members and the reference model is reduced: the reference production data are now located in the range given by the ensemble members. In addition, the final mean reservoir model defined from the final ensemble reproduces the behavior of the reference model better than the initial ensemble mean model. For the two wells, the estimation of water breakthrough during the history-matching period is improved significantly compared with the initial ensemble. More specifically, starting from initial ensemble members that do not reproduce the reference water cut data, the assimilation with EnKF improves water cut results at least for well PRO-15: the WCT is fairly well captured by the corrected ensemble through the history-matching interval. On the other hand, the WCT match is still poor for well PRO-5.

Matching WCT data, in general, is a difficult task because of its highly nonlinear dependence to model parameters. In terms of production responses, the corrected models provide reasonable predictions for BHP and SOR. As expected, the predicted WCTs are less reliable because the assimilation of the data during the history-matching phase led to smaller improvements. Overall, the updated ensemble outperforms the initial ensemble.

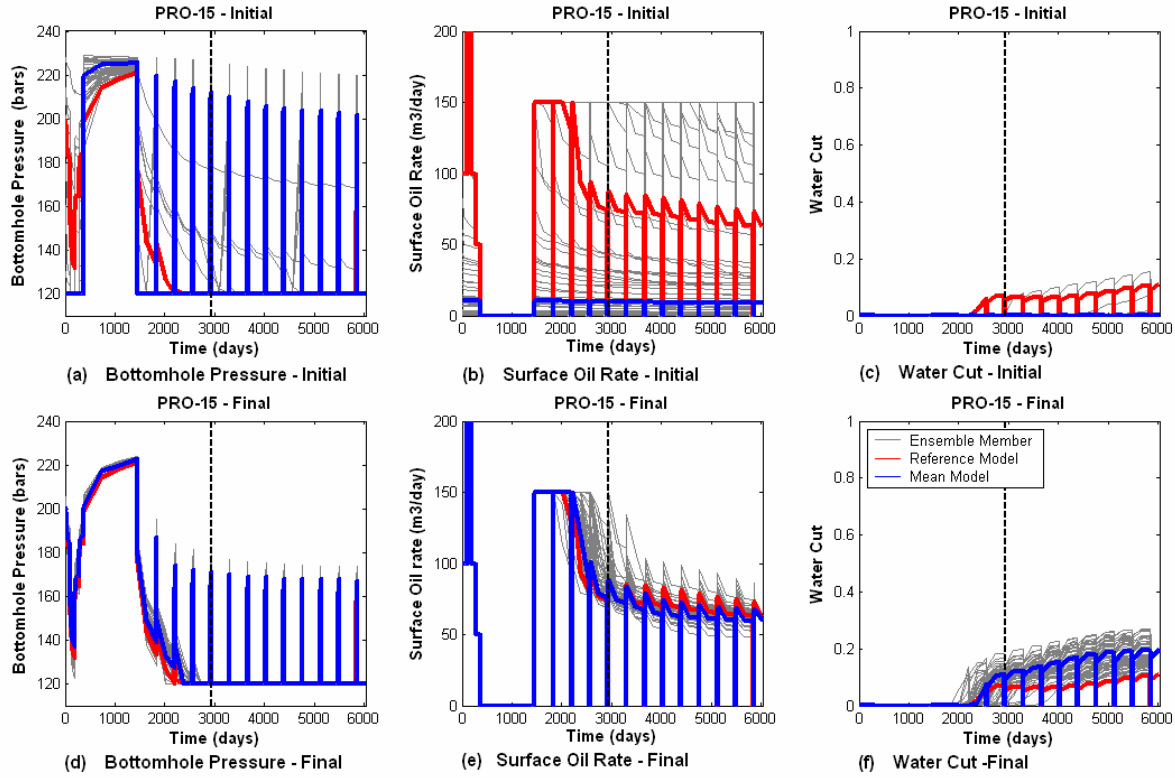


Figure 4-6 Production data simulated for well PRO-15 with an ensemble of size 50 during the history-matching and prediction periods. First row: Initial ensemble. Second row: Final ensemble updated with EnKF. The black dashed vertical line at 2936 days indicates the time limit between the history-matching interval (0-2936 days) and the prediction interval (2936-6025 days). The gray curves correspond to the ensemble members, the red curves to the reference model and the blue curves to the results obtained with the reservoir model computed as the mean of the ensemble.

As evidenced in Chapter 3, the choice of the initial ensemble can have an impact on the assimilation results, especially for a small ensemble size. To take this effect into account, four additional ensembles of 50 members were considered and updated with the same assimilation procedure.

To assess the results qualitatively, the $RMS_{prod,2}$ is calculated for the well production responses of the ensemble mean model at each assimilation step (equation 3.3).

Figure 4-7 represents the evolution of the RMS value with time for bottom hole pressure, surface oil rate and water cut. The values are normalized to the initial RMS value in each

case. The RMS for BHP and SOR show a decreasing trend (except for Ensemble 5), but the WCT RMS evolution is more chaotic. These results again stress the difficulty to match WCT compared to matching BHP and SOR. Furthermore, our experiments have shown that when the ensemble size is increased to 100 or more, the differences between initial ensembles in terms of RMS results become less significant.

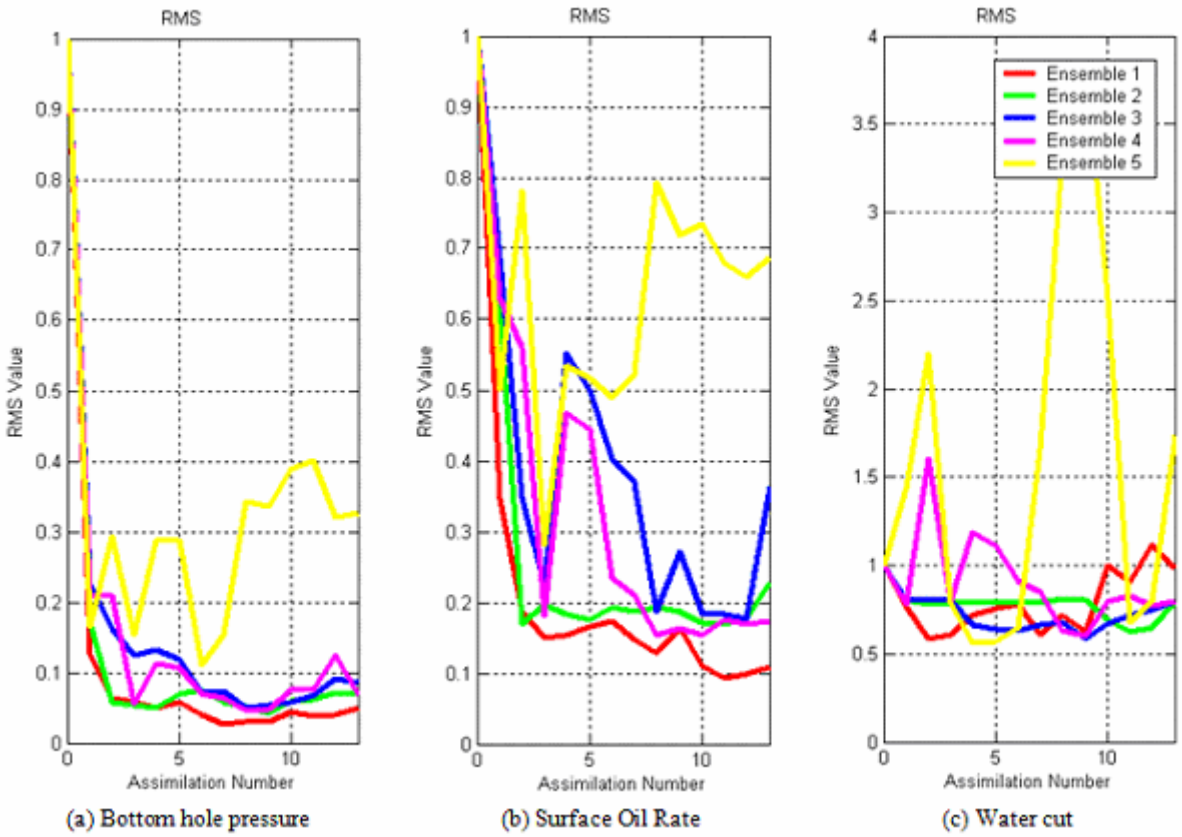


Figure 4-7 $RMS_{prod,2}$ of production data. (a) BHP (Bottom Hole Pressure), (b) SOR (Surface Oil Rate), (c) WCT (Water Cut) for 5 ensembles of size 50. The value are normalized on the initial value (t=0).

4.2.2 Results for petrophysical properties

Table 4-1 summarizes the properties of the porosity and permeability fields. Figure 4-8 and Figure 4-9 display, for Ensemble 1, the mean models derived from the initial porosity and permeability fields, respectively. They are almost uniform and equal to the porosity and permeability means, which are assumed to be stationary (Table 4-1).

By matching production data sequentially in time, heterogeneities and streaked features appear progressively in the porosity and permeability mean models from one assimilation time to the next. Figure 4-10 and Figure 4-11 show the mean corrected porosity and permeability models after performing 13 assimilations, *i.e.* 13 successive history-matching of

production data. The main features of the reference model (low and high porosity and permeability streaks in layers 1, 3 and 5 together with low and high porosity and permeability patches in layers 2 and 4) are retrieved in the corrected models. The porosity fields for layers 1, 3 and 5 were overestimated (the porosity values were expected to be lower than 0.35: they were trimmed when above).

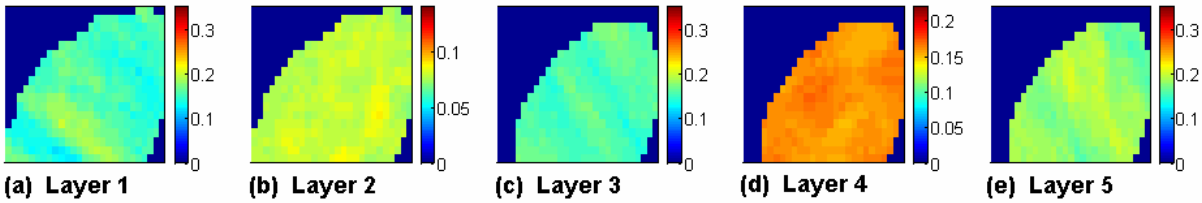


Figure 4-8 Mean of initial ensemble for porosity distribution in the 5 layers.

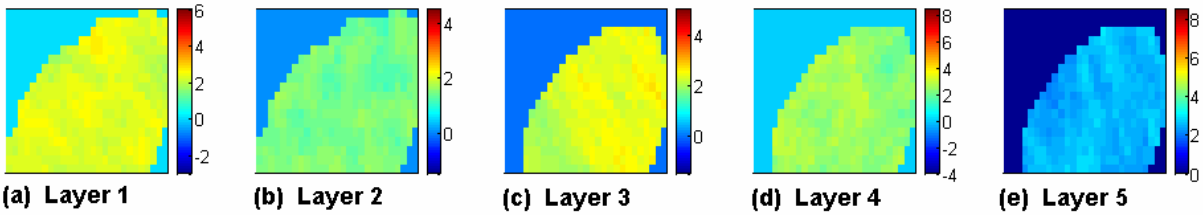


Figure 4-9 Mean of initial ensemble for permeability (ln k_h) distribution in the 5 layers.

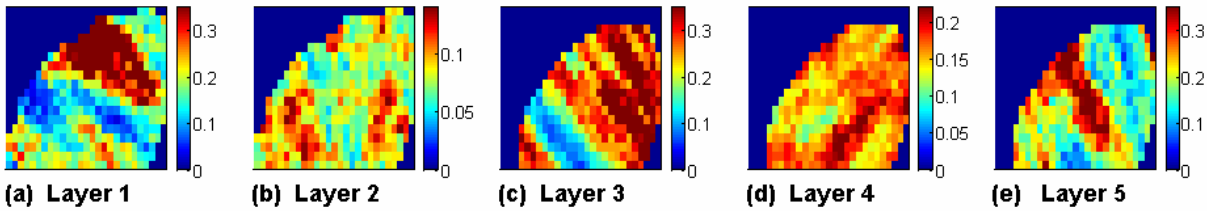


Figure 4-10 Mean of final ensemble for porosity distribution in the 5 layers.

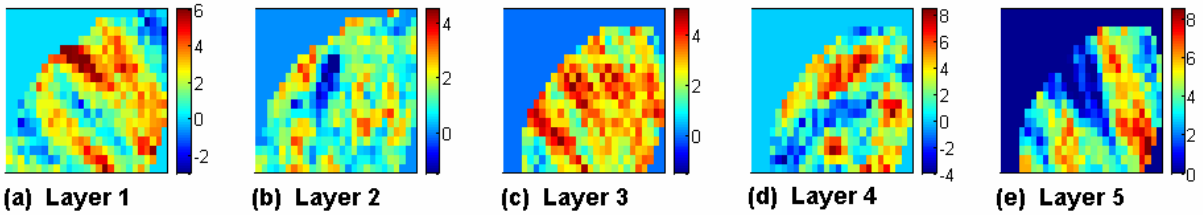


Figure 4-11 Mean of final ensemble for permeability (ln k_h) distribution in the 5 layers.

The overestimation is less significant for permeability fields: the updated fields are of the same order of magnitude as the reference model. Gao *et al.* (2006) related the overestimation problem to the constraints on well pressures and surface oil rates in the fluid flow simulation, and reduced it by removing the bounds on the well bottom-hole pressure. This problem is mentioned in other applications of the EnKF (*e.g.* Gu and Oliver, 2005). According to Devegowda *et al.* (2007), the overestimation is a result of the spurious correlations in the

covariance matrix due to the finite ensemble size. They proposed a mitigation strategy by covariance localization.

It can also be shown that the assimilation process induces a decrease in the variance of the porosity and permeability ensemble members, especially around well locations. As wells are perforated in layers 3, 4 and 5, variance reduction is more significant in these layers than in layers 1 and 2. Figure 4-12 and Figure 4-13 display, for Ensemble 1, the variance computed for the porosity and permeability models in layer 4. The first graph gives the variance for the initial ensemble, the middle one for the ensemble obtained after 7 assimilations and the last one for the final ensemble (13 assimilations). Although variability reduction is a natural consequence of data assimilation, the updated ensemble should be representative of the variability of model parameters. Hence, excessive variance reduction should be avoided. As already evidenced in Chapter 3, this can be ensured by using a larger ensemble. Moreover, the model noise is neglected in petroleum applications of the EnKF and its inclusion may help to mitigate this problem.

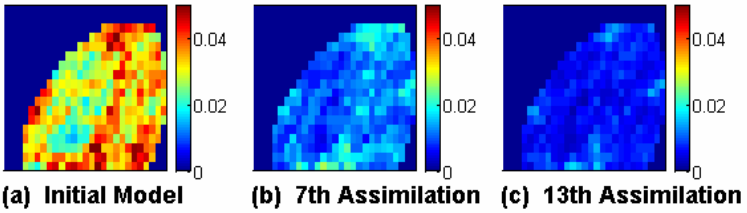


Figure 4-12 Variance of the ensemble for the porosity field in layer 4: (a) Initial ensemble, (b) Intermediate ensemble, (c) Final ensemble.

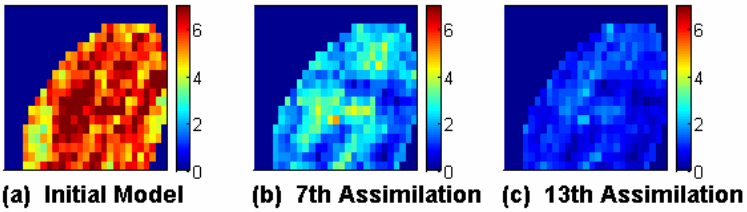


Figure 4-13 Variance of the ensemble for permeability field in layer 4: (a) Initial ensemble, (b) Intermediate ensemble, (c) Final ensemble

The distance between the assimilated porosity and permeability models for each ensemble member and the corresponding reference models is quantified by the $RMS_{stat,1}$ (equation 3.4).

This RMS value quantifies the convergence towards the reference reservoir parameters or static data. Figure 4-14 shows, for Ensemble 1, the static RMS values computed, at each assimilation, for porosity and horizontal permeability logarithm in the five layers (all values are normalized to the static RMS values computed at time 0). A decreasing trend is evident

for most of the parameters, except for porosity in layers 1, 3 and 5. As explained above, these layers control most of the flow and their updated porosities were overestimated. The effect of the initial ensemble was also studied on the RMS value: although the RMS values are not the same for all ensembles, the same trend was observed for different ensembles of size 50.

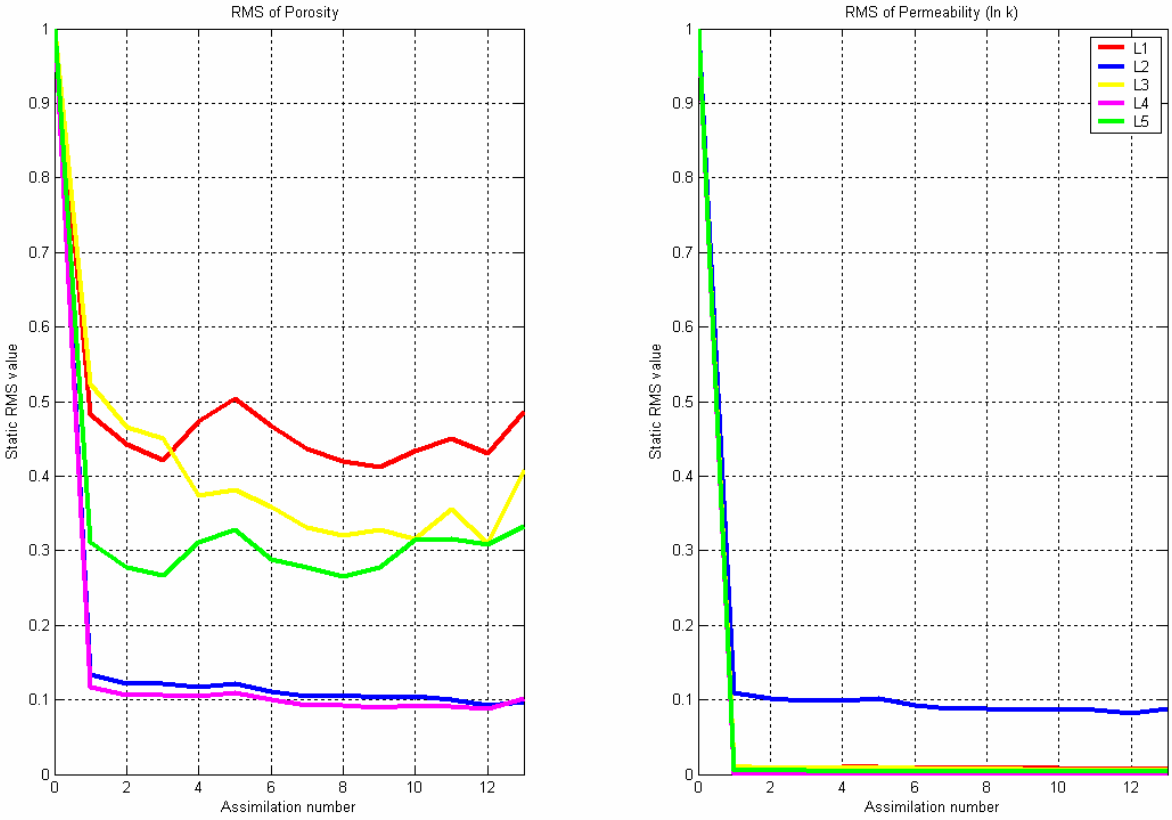


Figure 4-14 Static RMS for porosity and horizontal permeability log (ln kh) in layers 1 to 5. The values are normalized on the initial time (t=0).

4.3 Sensitivity analysis

Performance of the EnKF for history-matching and prediction depends on several parameters, among which ensemble size, measurement data uncertainty, assimilation time step are important. As in Chapter 3, a set of experiments illustrates the influence of these parameters on the results. In order to reduce the effect of the initial ensembles on the experiments, five different ensembles are generated for the target parameter and the same assimilation process is performed for each of them. The RMS values presented hereafter are the average RMS values over the five ensembles. The effect of the initial ensemble on predictions of the cumulative oil production is the same as the one presented in Chapter 3, except for the fact

that an ensemble of size 100 is more appropriate than 50. More details on this experiment are available in Heidari *et al.* (2010).

4.3.1 Ensemble size

As explained in section 3.2.1, increasing the ensemble size provides a better estimation for the error covariance matrix. Therefore, a better performance of the EnKF is expected in terms of regenerating the reference porosity, permeability and well production data. We apply here the EnKF considering ensembles of sizes 50, 100, 200 and 500. We consider the same data uncertainties as above. The assimilation intervals are indicated by the red dashed vertical lines in Figure 4-4.

Two metrics quantify the performance of the different ensembles: the static RMS ($RMS_{stat,2}$ in equation 3.5) and the dynamic RMS ($RMS_{prod,1}$ in equation 3.2).

Table 4-2 Effect of ensemble size – Average value of the dynamic $RMS_{prod,1}$ obtained with several ensembles of increasing size, normalized on the value obtained with an ensemble size of 50.

	Ensemble Size			
Property	**50**	**100**	**200**	**500**
BHP	1	0.33	0.27	0.25
SOR	1	0.29	0.19	0.17
WCT	1	0.44	0.44	0.41

Table 4-3 Effect of ensemble size – Average value of the static RMS obtained with several ensembles of increasing size, normalized on the value obtained with an ensemble size of 50.

	Ensemble Size			
Property	**50**	**100**	**200**	**500**
Phi-L1	1	0.62	0.55	0.54
phi-L2	1	0.74	0.68	0.66
Phi-L3	1	0.72	0.64	0.64
Phi-L4	1	0.67	0.64	0.65
Phi-L5	1	0.62	0.67	0.65
Ln kh-L1	1	0.67	0.66	0.61
Ln kh-L2	1	0.77	0.69	0.68
Ln kh-L3	1	0.83	0.69	0.72
Ln kh-L4	1	0.81	0.77	0.74
Ln kh-L5	1	0.71	0.71	0.65

Table 4-2 provides the $RMS_{prod,1}$ values determined for different ensemble sizes. The values are normalized to the average RMS obtained for the ensembles of size 50. At first glance, it appears that increasing the ensemble size from 50 to 100 strongly decreases the RMS value. A larger increase in the ensemble size has a less significant effect. Therefore, an ensemble of size 100 is more appropriate considering CPU costs. **Error! Not a valid bookmark self-**

reference. provides the $RMS_{stat,2}$ values for ensembles of increasing size. These values are normalized by the average RMS error determined for the ensembles of size 50. The larger the ensemble size, the lower the RMS error for most of the static properties. Furthermore, increasing the ensemble size from 200 to 500 reduces the static RMS error, but less than when passing from 50 to 100 or 200. Thus, an ensemble of size 100 to 200 is well-suited for the considered problem.

4.3.2 Data uncertainty

We propose now to assess the effect of measurement uncertainty on the performance of the EnKF. Under the Uncertainty 1 case, data uncertainties were set to 2 bars for BHP, 2 m^3/day for SOR and 0.02 for WCT. The Uncertainty 2 case assumes an error of 3 bars for BHP, 3 m^3/day for SOR and 0.03 for WCT. The quality of the results is again measured by the $RMS_{prod,1}$ error (equation 3.2). The results presented in Table 4-4 are normalized to the average RMS value obtained for the ensembles of size 50 and the "Uncertainty 1" setting to make the comparison of results easier. In the “Uncertainty 2” configuration, the RMS errors computed for the production data are increased in comparison with “Uncertainty 1” for a given ensemble size. Apart from the ensembles of size 50, our experiment demonstrates that even with a higher measurement uncertainty, the final ensembles outperform the initial ones.

Table 4-4 Effect of measurement uncertainty – Average value of the dynamic $RMS_{prod,1}$ obtained with two increasing levels of uncertainty and several ensembles of increasing size, normalized on the value obtained with an ensemble size of 50.

Uncertainty	Uncertainty 1			Uncertainty 2		
Ensemble Size	**50**	**100**	**200**	**50**	**100**	**200**
BHP	1	0.33	0.27	1.26	0.70	0.61
SOR	1	0.29	0.19	1.11	0.49	0.37
WCT	1	0.44	0.44	1.02	0.98	0.96

Haugen *et al.* (2006) have also performed a sensitivity test on measurement uncertainties for an application of the EnKF on a North Sea field case. They claimed that the choice of the measurement uncertainty is less important; assimilations with increased and decreased levels of uncertainty provided improved estimates compared to the initial reservoir models. Our experiments, in this chapter and Chapter 3, have resulted in the same conclusion.

4.3.3 Assimilation step

The time interval between two consecutive assimilations is also an important issue for practical EnKF applications. Smaller assimilation steps may be required to capture significant perturbations or nonlinearities in fluid flow. These perturbations can be induced by addition of a well or variations in flow rates.

The production scheme followed for this case study involves several changes in the flow rate. Therefore, we compare the performance of the EnKF for two distinct time step settings. The first assimilation time partitioning, “Time Step 1”, is reported in Figure 4-4. In the second partitioning, “Time Step 2”, each of the previous time steps is divided in two smaller steps. We considered ensembles of size 50, 100 and 200 for the Uncertainty 1 metric.

Results are compared using the $RMS_{prod,1}$ (equation 3.2). The RMS errors reported in Table 4-5 are normalized to the average RMS values determined for the ensembles of size 50 with the "Time Step 1" setting. Keeping the ensemble size constant while decreasing the time step size makes the RMS smaller for the target production responses. In addition, decreasing the time step size while increasing the ensemble size improves the match even more.

Table 4-5 Effect of assimilation interval – Average value of the dynamic $RMS_{prod,1}$ obtained with two decreasing sets of assimilation time intervals and several ensembles of increasing size, normalized on the value obtained with an ensemble size of 50.

Time Step	Time Step 1			Time Step 2		
Ensemble Size	**50**	**100**	**200**	**50**	**100**	**200**
BHP	1	0.33	0.27	0.60	0.28	0.24
SOR	1	0.29	0.19	0.56	0.21	0.15
WCT	1	0.44	0.44	0.50	0.35	0.29

4.4 Conclusions

In this chapter, we have applied the EnKF for history-matching of a rather complex synthetic case study, the PUNQ-S3. This application has provided us with a better understanding of the method on a practical point of view and highlighted additional problems compared to the 2D synthetic case considered in Chapter 3. In the first experiment, a fairly small ensemble of size 50 is used to provide the initial uncertainty in porosity and permeability. It was observed that by sequentially matching the production data (BHP, SOR and WCT), porosity and permeability fields were gradually adjusted so as to reproduce the reference production data. The water cut was shown to be quite difficult to match, possibly due to the highly nonlinear

dependence of this data to reservoir parameters. Moreover, the small size of the ensemble may have resulted in spurious correlations in the error covariance matrix leading to the overestimation of porosity in several layers.

We also performed a set of sensitivity tests to assess the role of several parameters on history-matching results such as the ensemble size, the data uncertainty, the assimilation time step. The following conclusions, in agreement with those of Chapter 3, can be drawn:

- increasing the ensemble size results in a better match for the production data (BHP, SOR and WCT), but there exists a trade off between the increase in computation time and the improvements in the results. Our experiments suggest that an ensemble size of 100 to 200 is appropriate to obtain acceptable results with EnKF. However, this number is case-dependent;
- uncertainty level in measured data is of less significance: data assimilation with EnKF improves model estimates whenever the size of the ensemble is large enough.
- decreasing the assimilation time step is required to better capture the abrupt changes in dynamics of the model.

5 Facies Model Case Study

Facies models are used to describe different rock types and depositional environments in a reservoir model. Therefore, in such a reservoir, there exists a finite number of facies types with specific petrophysical properties, *i.e.* porosity and permeability. Adjusting facies models in order to regenerate well production data is one of the most difficult tasks in history-matching due to the discontinuous and non-Gaussian nature of facies models. They cannot be directly updated by the EnKF. Several methods were proposed in the literature to circumvent this problem. They are first reviewed. Then, we apply the EnKF to a 2D synthetic reservoir populated by two facies.

5.1 The Ensemble Kalman Filter methodology for adjusting facies models

Mathematically speaking, facies models are considered as "indicators". Indicators are random functions that take discontinuous values from a predefined set of values. There are several simulation methods for generating such random functions (Chilès et Delfiner, 1999). The main difficulty with facies models is their non-continuous nature, which results in multimodal distributions. Besides, any application of the EnKF requires Gaussian random functions. Therefore, most temptative applications of the EnKF for adjusting facies models are based on the perturbation of underlying Gaussian fields. These Gaussian fields are considered as the static variables in the state matrix for each EnKF ensemble member. The new facies distributions after assimilation are obtained by transformation of the updated Gaussian field. An overview is given hereafter.

- Pluri-Gaussian Method (Liu and Oliver, 2005)

  In Liu and Oliver (2005), the Pluri-Gaussian method (Le Loc'h and Galli, 1997) is used to simulate the facies distribution. Two Gaussian random fields, $Z1$ and $Z2$, are defined on the whole reservoir model. Then, a threshold truncation map is used to assign facies at each location, depending on the values of $Z1-Z2$. These two Gaussian fields are used in the state matrix and updated at each assimilation step by the EnKF. The corresponding updated facies models are then obtained by applying the truncation map to the updated values of $Z1$ and $Z2$. More details about the

conditioning to facies observations are available in Liu and Oliver (2005). Two main disadvantages for this method are known. First, if the facies model is complex, it is not straightforward to determine an appropriate mapping and suitable structural properties for the underlying Gaussian fields. The second one is a common disadvantage for all the facies applications of the EnKF, regardless of the facies simulation method: the relation between facies models (defined by this methodology) and well production data is highly nonlinear. Therefore, iterative approaches are required to match well production data.

- Level Set Method (Moreno and Aanonsen, 2007)

For reservoir facies modeling, level set functions can be used to define the boundaries between different facies types. According to the number of facies types, several level set functions may be required. In this method, facies boundaries can be moved according to a velocity function. This velocity function is the key point in adjusting the facies models to regenerate the well production data. In the method proposed by Moreno and Aanonsen (2007), the velocity function is modeled as a Gaussian random field and is updated by the EnKF method at each assimilation step. The method was applied to a 2D synthetic reservoir model and was capable of regenerating most features of the reference facies model. Moreover, facies proportions were correctly estimated by the updated models. The only disadvantage is the severe complexity of the method for sophisticated facies models.

- Gaussian Mixture Models (Dovera and Della Rossa, 2007)

Dovera and Della Rossa (2007) modified the EnKF method to make it appropriate for Gaussian Mixture models. Gaussian mixture models are used to handle multimodal distributions. They first apply this new algorithm to a mathematical case study with a bimodal distribution. This modified version succeeded in regenerating this distribution. However, a reservoir engineering model is not still tested.

- Kernel Methods (Vabø *et al.*, 2008)

In order to deal with the non-Gaussian and multimodal nature of facies distribution, Vabø *et al.* (2008) proposed to map the facies models to a higher order feature space. This method can be considered as a parameterization method. By making a proper mapping, facies models can be represented by Gaussian variables in the feature space. When using kernel methods, the mapping is never explicitly performed. The update of the EnKF is performed in the feature space and the updated values are then back-transformed. However, calculating the back transform is not straightforward. The method is claimed to be as efficient as the traditional EnKF in terms of computations. It is also said that it easily applies to large scale reservoir

models. Vabø *et al.* (2008) applied this method to a real field case and were capable of regenerating the observed well production data to a reasonable degree. However, they observed that the final facies models were rather homogeneous and did not reproduce the geological features of the actual facies model.

As our facies case study is a simple one with only two facies types, we make use of the "Truncated Gaussian Simulation" method and we aim to combine it with the EnKF.

5.2 Truncated Gaussian Simulation

We consider a reservoir model with m facies types or m indicators. One way to obtain these indicators is to apply $m-1$ thresholds to a standard Gaussian random function and assign each of the m zones to one of the indicators. This idea is the key of the truncated Gaussian method. It is depicted in Figure 5-1.

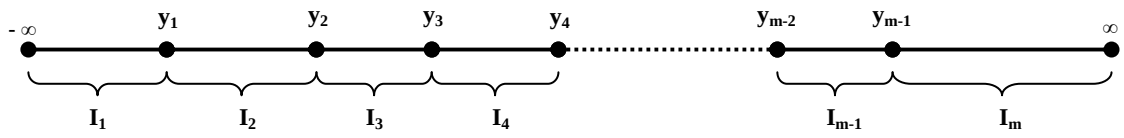


Figure 5-1 Truncated Gaussian Simulation Method. y_i is the corresponding threshold for indicator I_i.

The Gaussian random function $Y(x)$ takes values in $\Re$. Functions $I_i(x), i=1,...,m$ represent each of the indicators defined by the thresholds $y_i, i=1,...,m$ according to the following formula:

$$I_i(x) = 1_{y_{i-1} \le Y(x) < y_i} \qquad -\infty = y_0 < y_1 < ... < y_{i-1} < y_i < ... < y_m = +\infty \tag{5.1.}$$

Thresholds y_i are chosen so as to reproduce facies proportions p_i:

$$y_i = G^{-1}\left(\sum_{j=1}^{i} p_j\right) \ , \ i=1,...,m-1 \tag{5.2.}$$

where G is the Normal standard cumulative density function. Figure 5-2 represents the mapping between the indicators, Gaussian thresholds and proportions.

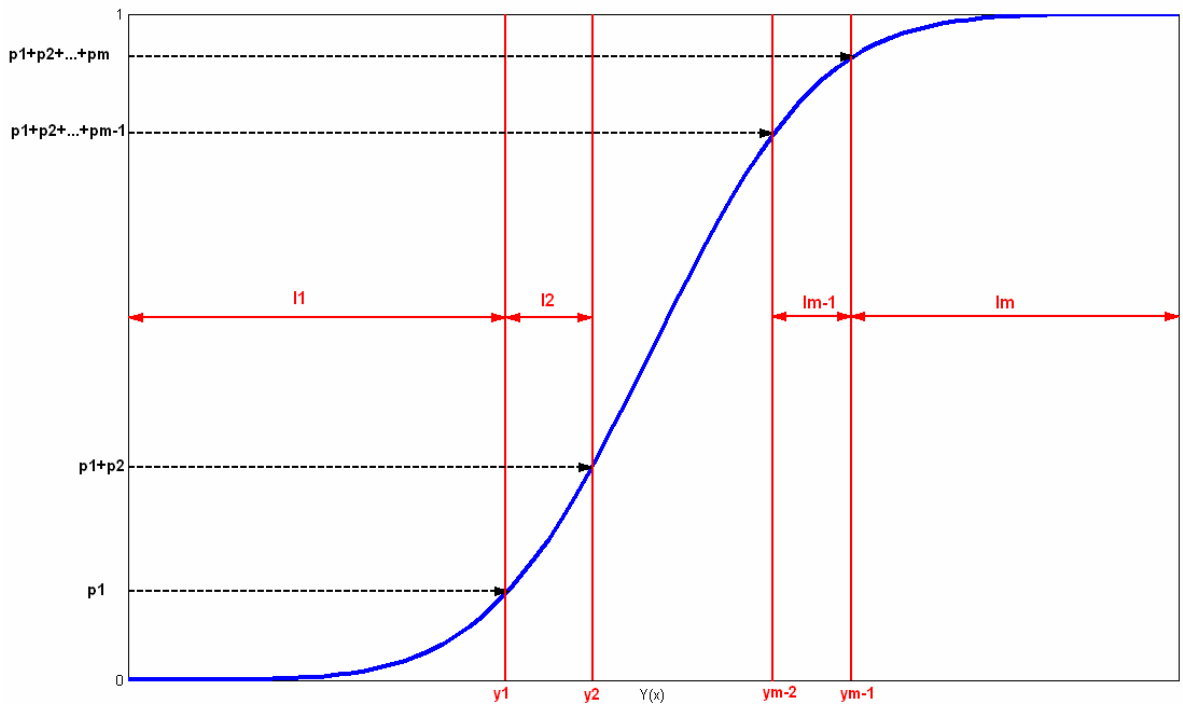


Figure 5-2 Schematic on the Truncated Gaussian Method

Starting from the first indicator I_1 and its proportion p_1, we estimate threshold y_1: the set of Gaussian values smaller or equal to y_1 are attributed to indicator I_1. A similar approach is followed for the other indicators.

The truncated Gaussian method is appropriate for simple facies models. Otherwise we can use its extension that is the pluri-Gaussian method. Then, when adjusting facies models with the EnKF, the Gaussian random function Y is considered as the static variable to be updated.

The truncation is applied to determine the facies distributions.

5.3 Two dimensional facies case study

5.3.1 Test case description

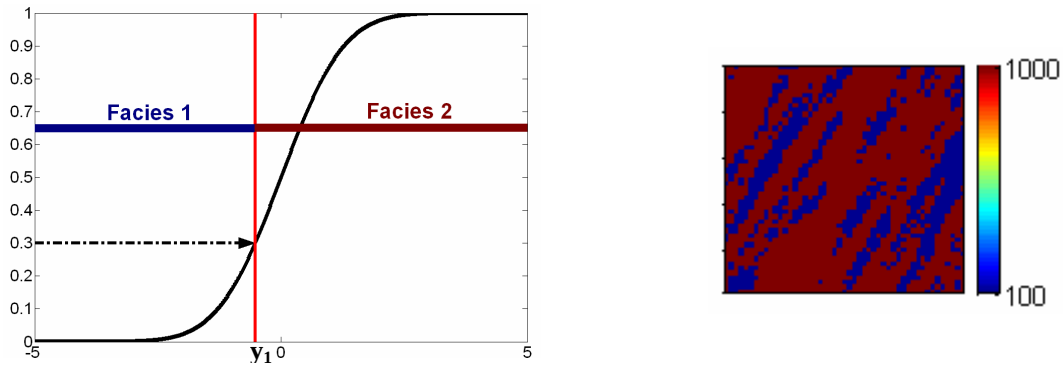
The numerical facies model for this case study is built over a 50×50 grid. The reservoir is populated with two facies. They are assigned constant petrophysical properties and proportions (Figure 5-3). The porosity is set to 0.18 all over the reservoir.

Table 5-1 Facies proportions and properties

Facies	Proportion	Permeability

1	30 %	100 mD
2	70 %	1000 mD

The variogram for the underlying Gaussian random function is spherical with the main anisotropy direction defined by an angle of 60 degrees with respect to the X-axis. The horizontal variogram ranges are 300 m and 40 m. The FFT-MA algorithm (Le Ravalec *et al.*, 2000) is used to generate this standard Gaussian field. Then, we refer to the mapping described in Figure 5-3 (a) to assign permeability values and facies type to each grid cell. According to this mapping, the value y_1 of the Gaussian random function that corresponds to a cumulative probability of 0.3 is first determined (this process is represented by the black dashed arrow). Then, grid cells for which the Gaussian random field is lower than y_1 are associated to facies 1 whose permeability is 100 mD while the others are associated to facies 2 whose permeability is 1000 mD. The reference permeability field is represented in Figure 5-3 (b).

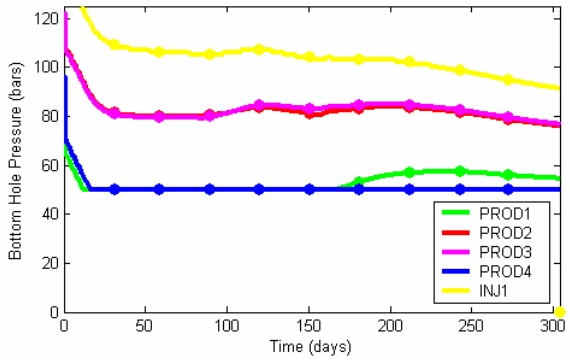


(a) Mapping model for facies models (b) Reference permeability field (mD)

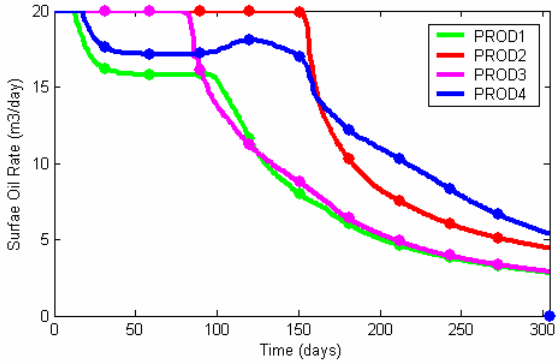
Figure 5-3 (a) Mapping model for facies proportions, (b) Reference permeability field for the facies case study. The dashed black arrow in (a) represents the facies 1 proportion, *i.e.* 0.3. The red vertical line represents its corresponding Gaussian random function value. In figure (b), facies 1 corresponds to blue values (with a permeability of 100 mD) and facies 2 to red values (with a permeability of 1000 mD).

Five wells are located in the reservoir (4 producers and 1 injector) with the same configuration as in Chapter 3 (Figure 3-1). The initial reservoir pressure is 125 bars. The production planning consists of water injection as soon as production starts at day 0. All producers produce with a rate of 20 m^3/day. The flow rate constraint is changed to a bottom hole pressure (BHP) control when BHP drops below 50 bars. The primary constraint for the injection well is the maximum injection rate, which is set to 80 m^3/day. The secondary constraint is a maximum injection pressure of 300 bars. The STONE1 model was used for defining relative permeability curves with the associated end points and water capillary pressure tables. The initial water saturation is set to 0.15 and the residual oil saturation to 0.2. The compressibilities of water and oil are set to 0.10×10^{-4} bar^{-1} and 0.10×10^{-3} bar^{-1},

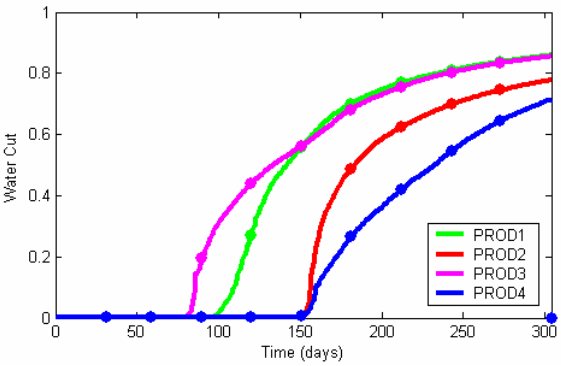
respectively. The corresponding densities are 1.05 g/cm^3 and 0.9 g/cm^3. The industrial IFP reservoir simulator PumaFlow^{TM} is used for reservoir simulation with a Black Oil PVT type. Bottom Hole Pressure (BHP), Surface Oil Rate (SOR) and Water Cut (WCT) were simulated for the reference permeability field over a period of 304 days (10 months). They are displayed in Figure 5-4.



(a) Bottom hole pressure for the reference case



(b) Surface oil rate for the reference case



(c) Water cut for the reference case

Figure 5-4 Reference values for (a) Bottom hole pressures (b) Surface oil rates (c) Water cuts. The solid dots indicate the timing at which production data will be assimilated with the EnKF.

5.3.2 Assimilation procedure

Our current purpose is the same as the one in Chapter 3. The reference permeability field (Figure 5-3 (b)) is considered as unknown. The reference well production data only are assumed to be known (Figure 5-4). Then, we try to identify a set of permeability (or facies) realizations that regenerate both the features of the reference permeability (or facies) field and the reference well production data. The production data are assimilated at the beginning of each month up to 304 days (10 months). The assimilation times are represented by dots in Figure 5-4. The distributions of measurement errors are assumed to be Gaussian with standard deviations of 5% of the values measured for BHP, SOR and WCT.

The EnKF ensemble size is set to 50, *i.e.*, 50 different standard Gaussian random realizations are generated with the same variogram as the reference field. Then, the mapping described in Figure 5-3 (a) is applied to each of them to determine the corresponding permeability (or facies) realizations. Other parameters (porosity = 0.18, relative permeability curves, initial pressure = 125 bars and initial water saturation = 0.15) are kept constant without any uncertainty.

The state vector for each ensemble member j then includes:

- The underlying Gaussian random realization (static variable): $Y_{j,i}$, $i = 1,...,2500$;
- Pressure at all grid cells (dynamic variable): $P_{j,i}$, $i = 1,...,2500$;
- Water saturation at all grid cells (dynamic variable): $Sw_{j,i}$, $i = 1,...,2500$;
- Bottom hole pressure at injection and production wells: $BHP_{j,k}$, $k = 1,...,5$;
- Surface oil rate at production wells: $SOR_{j,k}$, $k = 1,...,4$;
- Water cut at production wells: $WCT_{j,k}$, $k = 1,...,4$.

Therefore, the state vector for the j^{th} ensemble member is defined as:

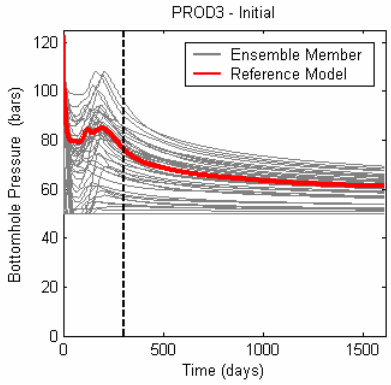
$$\psi_j = [Y_{j,1}, Y_{j,2}, ..., Y_{j,2500}, P_{j,1}, P_{j,2}, ..., P_{j,2500}, Sw_{j,1}, Sw_{j,2}, ..., Sw_{j,2500}, BHP_{j,1}, ..., BHP_{j,5}, SOR_{j,1}, ..., SOR_{j,4}, WCT_{j,1}, ..., WCT_{j,4}] \quad 5.1.$$

After performing the update, the mapping is applied to the updated field Y to estimate the updated set of permeability (or facies) realizations to be used at the next time step.

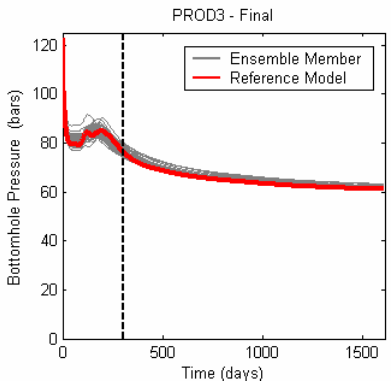
The permeability models obtained at the end of the assimilation period are used to perform predictions. They are taken back to day 0 and used to simulate the fluid flow behavior in the reservoir during the history-matching period (0-304 days) and the prediction period (305-1612 days).

5.3.3 History-matched production data and permeability (or facies) model

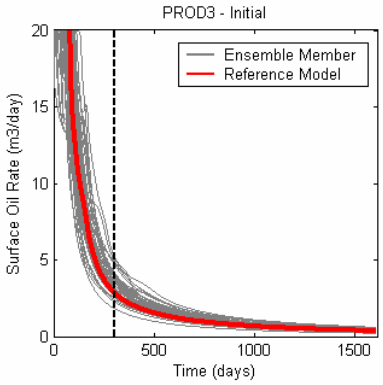
Production responses simulated with the initial and final ensemble members are compared in Figure 5-5 for well PROD3. The same type of results is obtained for other wells.



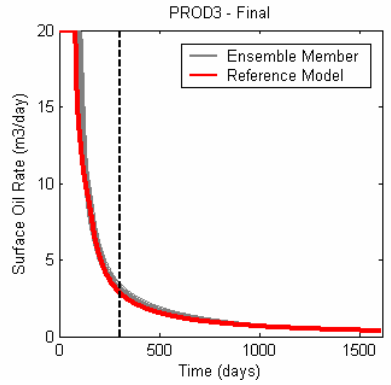
(a) Bottom hole pressure in PROD3 – Initial Ensemble



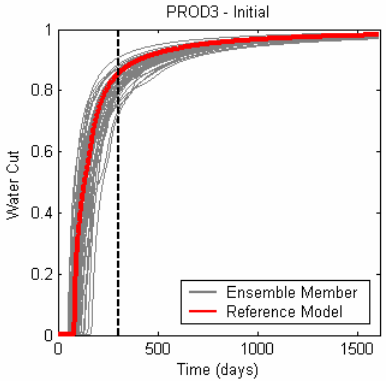
(d) Bottom hole pressure in PROD3 – Final Ensemble



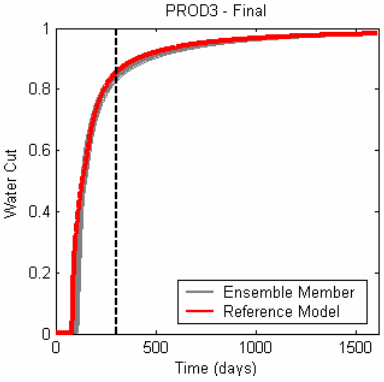
(b) Surface oil rate in PROD3 – Initial Ensemble



(e) Surface oil rate in PROD3 – Final Ensemble



(c) Water cut in PROD3 – Initial Ensemble



(f) Water cut in PROD3 – Final Ensemble

Figure 5-5 Production data simulated at well PROD3 with the 50 facies realizations of the initial ensemble (first column) and the final ensemble (second column). The black dashed line at 304 days indicates the time limit between the matching period (0-304 days) and the forecasting one (305-1612 days). The gray curves correspond to the ensemble members and the red one to the reference model.

Figure 5-5 (a) shows that the bottom hole pressure for the initial ensemble members is widely spread around the reference value. However, after going through the assimilation process (Figure 5-5 (d)), this spread reduces strongly and all final ensemble members follow closely the trend of the reference model in both history matching and prediction periods. The same type of results can be observed for surface oil rate (Figure 5-5 (b) and Figure 5-5 (e)) and water cut (Figure 5-5 (c) and Figure 5-5 (f)). In each case, the assimilations yielded final

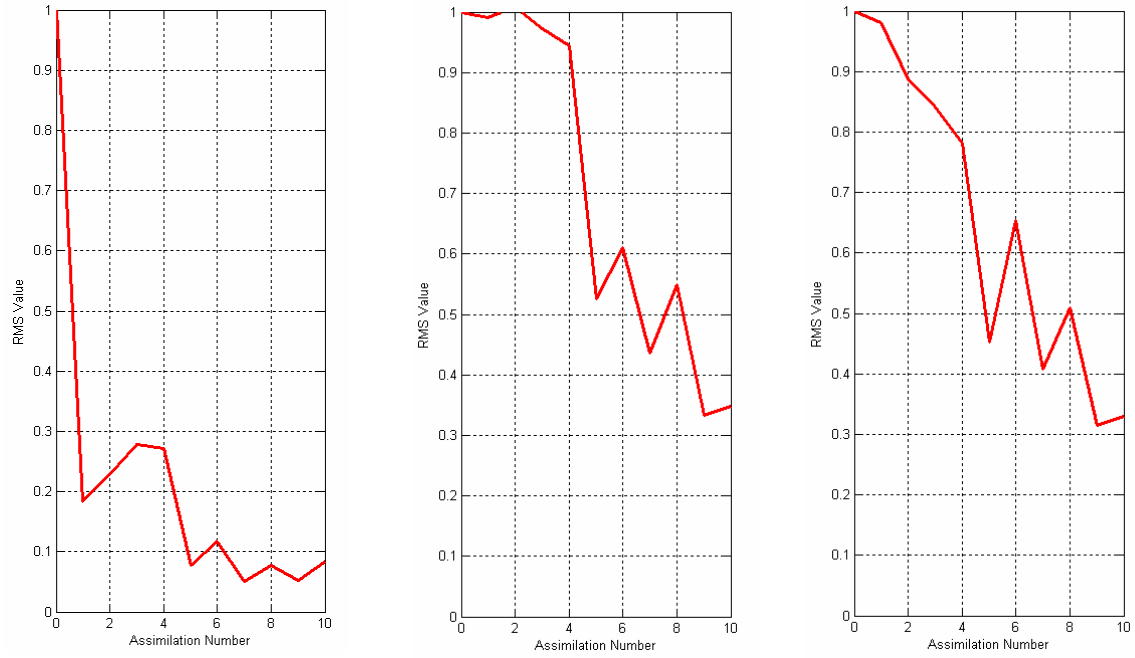
updated models that keep track of the reference model better than the initial members. However, the variability within the final ensemble members is considerably decreased. Therefore, the variation in production responses in the prediction phase is small.

These results are confirmed quantitatively by the Root Mean Square error defined in equation 3.2 ($RMS_{prod,1}$). This error calculates an average difference between the production responses for each ensemble member and the reference data. Performing the assimilations reduced the final RMS to 14%, 61% and 50% of its initial value for bottom hole pressure, surface oil rate and water cut, respectively (Table 5-2).

Table 5-2 $RMS_{prod,1}$ for the initial and final ensembles.

Dynamic RMS	BHP	SOR	WCT
Initial	2.6749	0.3745	0.0238
Final	0.3679	0.2277	0.0118
Final / Initial	0.1375	0.6080	0.4958

The performance of the assimilation procedure in regenerating the reference production data by the updated models can be also investigated by the RMS error defined for the ensemble mean model at each time step, $RMS_{prod,2}$ (equation 3.3).



(a) Bottom hole pressure (bars) **(b) Surface oil rate (m3/day)** **(c) Water cut**

Figure 5-6 $RMS_{prod,2}$ computed for (a) bottom hole pressure, (b) surface oil rate, (c) water cut throughout the assimilation procedure for the facies model case study.

It is plotted for bottom hole pressure, surface oil rate and water cut in Figure 5-6. A general decreasing trend is observed with minor oscillations. There may exist different reasons for

explaining these oscillations, the most probable one being the nonlinear dependence of the EnKF static state variables, *i.e.* the underlying Gaussian values, with the well production data. In addition, flow changes such as water breakthrough may have an effect (see Section 3.1.2).

Figure 5-7 represents four of the initial and final facies ensemble members for this assimilation scenario. The initial ensemble models were generated with the same simulation scenario, and a good level of variability can be observed between them. After performing the assimilations, the updated ensemble facies models exhibited similar features. Fortunately, most of these features are common with the reference facies model in Figure 5-3 (b). However, this may not be observed for a more complicated case study (Liu and Oliver, 2005).

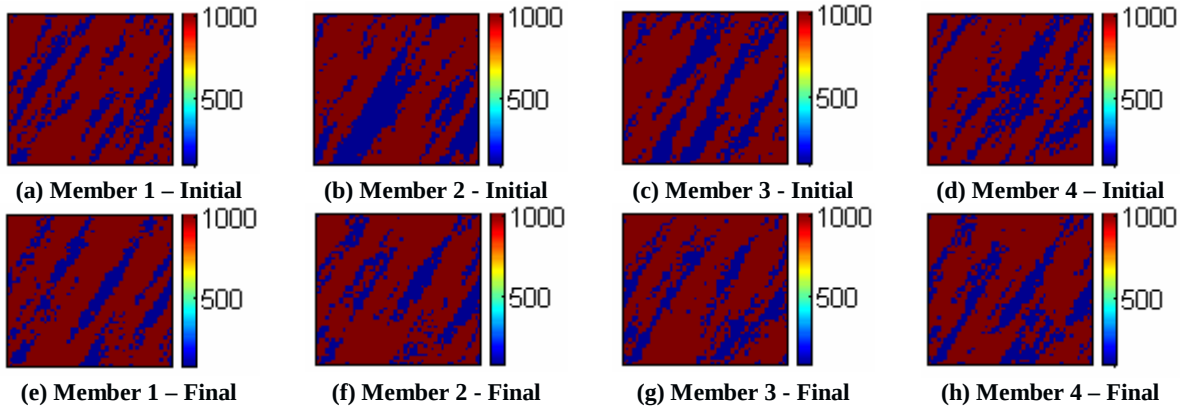


Figure 5-7 Four initial and final ensemble members for permeability. Facies 1 correspond to blue values (with a permeability of 100 mD) and facies 2 to red values (with a permeability of 1000 mD).

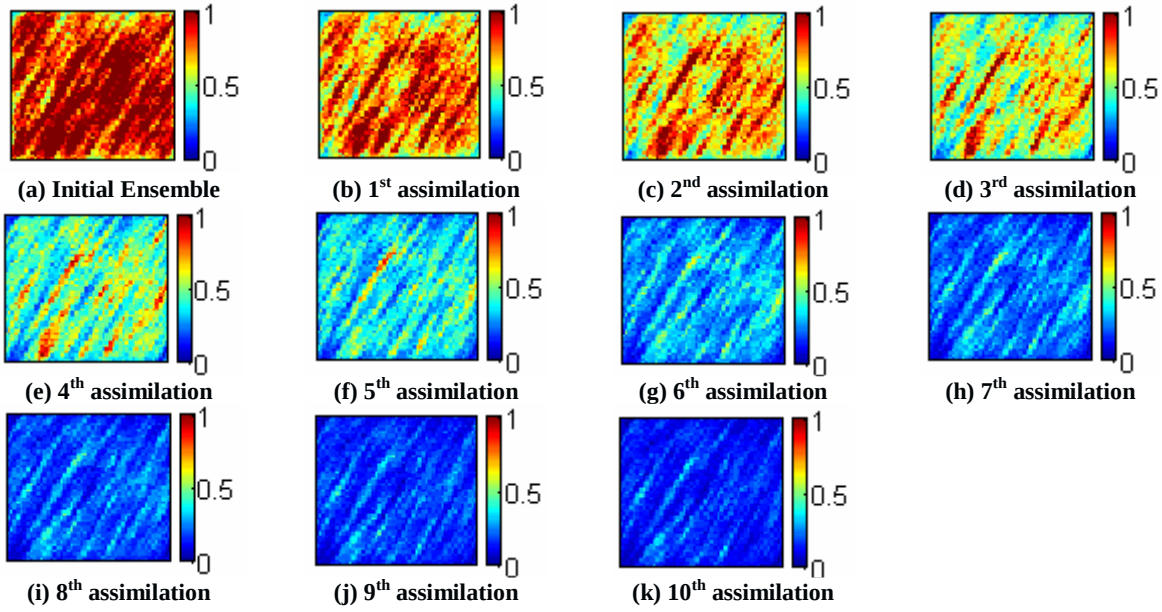


Figure 5-8 Underlying Gaussian field evolution: (a) Variance of the initial ensemble, (b)-(k) Variance of the updated ensemble against assimilation steps.

By performing the assimilations, the variability between ensemble members reduces. This can be shown by the decrease in the variance of the underlying Gaussian fields as shown in Figure

5-8. The variance of this ensemble at each assimilation step provides an idea about the subspace spanned by the ensemble members. As a direct consequence of assimilation, the variance has been reduced throughout time.

During the assimilation process, the underlying Gaussian fields stay adequately Gaussian, but their mean and variance undergo some changes. Therefore, the truncation procedure is not capable of preserving the facies proportions provided in Table 5-1. During the assimilations, the proportion of facies 1 is changing according to Table 5-3.

Table 5-3 Facies 1 proportions during assimilations

Time Step	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Proportion	0.2956	0.2851	0.2802	0.2711	0.2600	0.2770	0.2578	0.2622	0.2544	0.2560	0.2470

5.4 Conclusions

In this chapter, we used of the EnKF method for adjusting facies models in order to regenerate well production data. The facies models are not continuous and do not obey Gaussian distributions. Therefore, an intermediate methodology is required to be able to apply the EnKF to facies models. The "Truncated Gaussian Simulation" method was used to generate facies models from a set of underlying standard Gaussian fields considering a 2D reservoir. These Gaussian fields are considered as the EnKF static variables. After assimilation, the reference well production data are quite well matched and the final ensembles members capture several features of the reference facies model.

However, this algorithm may not be appropriate for more complex facies models. In addition, if facies observations are available at well locations, iterations may be required to ensure correct facies at these locations. One solution would be to update proportions spatially rather than the Gaussian fields in the EnKF setting.

II. Ensemble Kalman Filter and Parameterization Methods

Numerical models for real reservoirs consist of millions of grid cells, each populated with petrophysical properties such as porosity and permeability, saturation, *etc*. Although upscaling the reservoir model can strongly reduce this number, the resulting coarse reservoir grid still includes a large number of cells. Variational history-matching methods are usually not able to handle large numbers of parameters. Also, the perturbation of petrophysical properties in grid blocks must be done carefully to make sure that the perturbed model remains consistent with the prior information inferred from data. Therefore, parameterization methods have been developed to vary the geological model from a few parameters only while preserving the spatial variability of the model. Among them are the pilot point method (de Marsily, 1984), the gradual deformation method (Hu, 2000) and the probability perturbation method (Caers, 2003).

In this section, two of these parameterization methods, the pilot point and the gradual deformation methods, are considered to be combined with the EnKF. The main objective is to be able to prevent the EnKF updated models to depart from the spatial variability of the initial models. This may prevent the parameter overestimation problem encountered with the EnKF. To do so, we apply the EnKF to adjust a small number of parameters representing the static variables in the studied model: these parameters drive the update of the whole geological model. However, the nonlinear dependence between the measurements and the model variables (static and dynamic) impedes the performance of the EnKF. Any parameterization method contributes to strengthen this nonlinearity. In order to assess the performance of the EnKF when combined with these parameterization methods, it was applied to the 2D synthetic case study introduced in Chapter 3. The two variants of the EnKF proposed to integrate both the pilot point and the gradual deformation methods are detailed in Chapters 6 and 7.

6 Pilot Point Method (PPM)

The pilot point method was first introduced by de Marsily (1984). This parameterization method makes it possible to perturb the realization of a given property while preserving its spatial variability. The parameters are the values of the property of interest in some grid cells that are referred to as "pilot points". Any perturbation of these values is propagated to the whole realization through kriging. If the pilot points are sufficiently far from one another, the spatial variability defined by the variogram model is preserved. The petrophysical property values at pilot point locations can be considered as unknowns in history-matching. Thus, they are adjusted automatically in the optimization loop.

In this chapter, we introduce the pilot point methodology and propose an algorithm to combine the EnKF with this parameterization. The resulting method is then applied to the 2D case study used in Chapter 3. The influence of the number and locations of pilot points will be investigated.

6.1 Pilot point method

The pilot point method was first considered as an estimation method in the work by de Marsily (1984) about the interpretation of interference test data. The purpose was to generate a smooth permeability realization applying kriging to the permeability values assigned to the pilot points. La Venue and Pickens (1992) modified the existing algorithm to calculate pilot point locations automatically. The pilot point method was further extended to condition permeability realizations to pressure data in the work of RamaRao *et al.* (1995) by minimizing an objective (or cost) function consisting only of the mismatch term between the simulated and observed pressures (second term on the right hand side of equation 2.4). In addition, they proposed an algorithm to place the pilot points automatically in order to reduce the pressure data mismatch. The same method was applied to assess the probability of containment of radioactive waste at the Waste Isolation Pilot Plant (WIPP) site in New Mexico by La Venue *et al.* (1995). Later, the "Master Point" method was proposed by Gómez-Hernández *et al.* (1997): these authors used it to condition permeability realizations to production data. It was

at first differentiated from the pilot point method, but the work by RamaRao *et al.* (1995) et La Venue *et al.* (1995) showed their close resemblance.

The principle of the pilot point method is the following (see also Figure 6-1). Let us consider an unconditional y realization of some petrophysical property and *n* pilot points. Then, y can be conditioned to new values at pilot point locations by conditional kriging (Chilès and Delfiner, 1999):

$$y_c(\mathrm{u}) = y_{dK}(\mathrm{u}) + [y(\mathrm{u}) - y_K(\mathrm{u})] \qquad 6.1.$$

Vector u indicates location. y_{dK} is the kriging estimate built from the available hard data and values attributed to the pilot points, and y_K the kriging estimate computed from the values of y at pilot point locations. The so-obtained conditional realization y_c honors the pilot point values at pilot point locations.

The pilot point method can be used as a parameterization technique in a history-matching process. In this case, the pilot point values are considered as unknown and are adjusted through the minimization process. This method can result in numerical artifacts. First, the minimization of the objective function can lead to extremely low or high pilot point values. Therefore, RamaRao *et al.* (1995) suggested to bracket pilot point variations to circumvent this problem. The second limitation is about the spatial variability preservation: as the values of the pilot points are optimized independently, the spatial correlations are not preserved unless the points are separated by at least one correlation length. The number and the locations of the pilot points are thus subject to limitations. McLaughlin and Townley (1996) proposed to include a term into the objective function that penalizes the departure from prior information. This makes it possible to preserve spatial correlations between pilot points and to prevent pilot points from being assigned extreme (low and high) values. It should be noted that the same solution can be followed for the master point method (Gómez-Hernández *et al.*, 1997). Finally, Le Ravalec-Dupin and Hu (2007) proposed to perturb the pilot points' values on the basis of the gradual deformation. This approach also allows for avoiding extreme values and preserving spatial variability.

Another issue with the pilot points' method is whether all the pilot points should be incorporated simultaneously or sequentially in the optimization procedure. The simultaneous incorporation can result in irregular variations in the objective function values as noted by Bissel *et al.* (1997) and Xue and Datta-Gupta (1997). This drawback was not encountered when the points are accounted for sequentially. In addition, the work by La Venue *et al.* (1995) stressed that the sequential approach is more efficient.

As mentioned briefly, several automatic algorithms for pilot point placements were proposed in the literature. These methods involve sensitivity criteria based upon numerical gradients to optimally locate the pilot points. More details can be found in Cuypers *et al.* (1998) and Mezghani *et al.* (1998) for instance.

6.2 Algorithm combining the EnKF and the PPM

We propose in this section an algorithm combining both the EnKF and pilot point methods: the purpose is to better preserve the spatial variability of reservoir porosity and permeability distributions and to decrease the number of parameters.

As stated in Chapter 2, the EnKF describes the state of any model of interest at each time step with a set of static variables, dynamic variables and measurements. Here, the key idea is to consider only the values of pilot points as static variables in the state matrix. The general workflow is recapped in Figure 6-2.

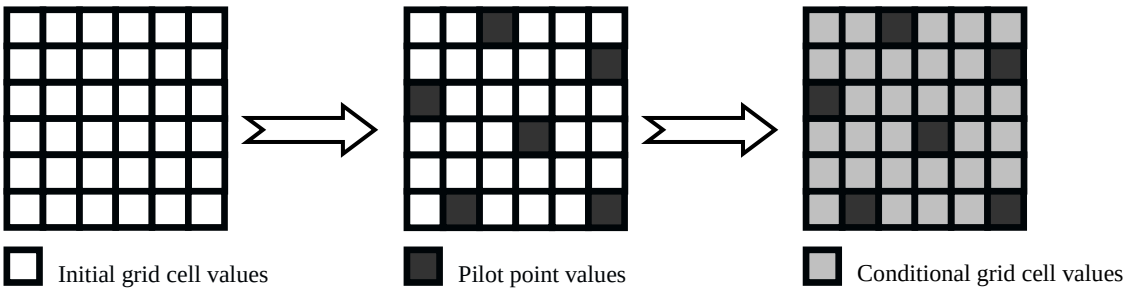


Figure 6-1 Pilot point method description

Let us consider an initial ensemble of size *N*, called "background ensemble" in the remaining of this chapter. It consists of *N* realizations for each property of interest generated using the prior knowledge inferred from the static data. Then, pilot point locations are fixed, based on some "intuition" about the problem at hand. In the present work, we did not make use of the existing automatic pilot point placement techniques. Then, at each assimilation time, the method can be split into two steps:

1. Forecast

- Condition each realization of the background ensemble to the pilot point values obtained from the update step at the previous assimilation. The resulting ensemble is called "Fluid flow ensemble" or "EnKF ensemble".

- Perform the forecast from the fluid flow ensemble and the updated dynamic variables from the previous assimilation. It results in a forecasted model containing the pilot point values, the forecasted dynamic variables and the simulated production data.

2. Update

- Apply the EnKF update step to the forecasted model.
- The resulting set of updated pilot point values and dynamic data are used as initial values for the forecast at the next time step.

Let us note that the pilot point values used at the first assimilation in the forecast are extracted from the background ensemble. The dynamic variables are given by the initial equilibrium in the reservoir models.

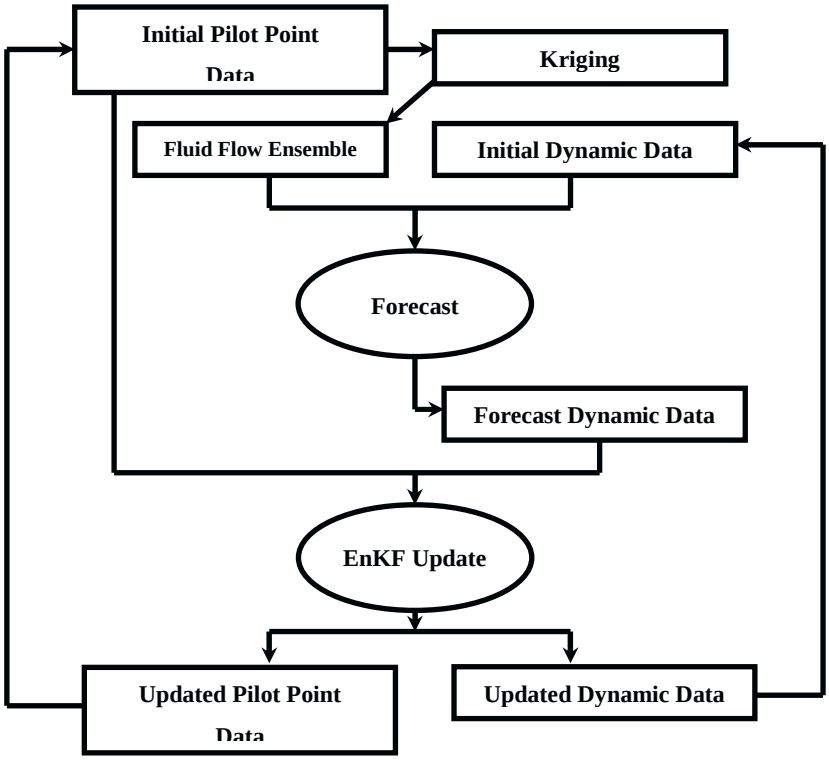


Figure 6-2 Workflow for the algorithm combining the EnKF with the pilot point method

6.3 Two dimensional case study

We applied the algorithm developed above to the two dimensional synthetic case study described in Chapter 3. The aim of this experiment is to obtain a set of permeability models conditioned to well production data. The production and assimilation criteria, assimilation time interval and measurement uncertainty level are the ones of the base case (see Chapter 3).

6.3.1 History-matching process

The 2D reservoir is modeled on a 50 by 50 grid. The dimension of each cell is 10m by 10m. As a first experiment, the pilot points are located at well locations, *i.e.*, the four corners and the middle point (see Figure 6-3).

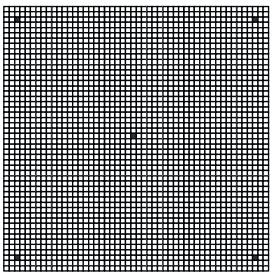


Figure 6-3 2D synthetic reservoir model. Pilot points locations are represented by the dark gray cells.

As detailed in Chapter 3, the reference permeability model is characterized by a spherical variogram. The angle between the main axis direction and X axis is 30 degrees. The main correlation lengths are 200m along the main axis and 40m in the perpendicular direction. Based upon this setting, the minimum distance between pilot points is 320m, which is larger than correlation lengths. Therefore, this configuration ensures the preservation of spatial variability.

We consider an EnKF ensemble with 50 models: 50 permeability realizations are generated with the FFT-MA algorithm (Le Ravalec *et al.*, 2000) following the specifications provided in Chapter 3. These realizations form our background ensemble. The values at pilot point locations are extracted from this set of realizations, hence providing the initial ensemble for the static data. The reservoir produces from equilibrium state: the initial pressure and saturation are known without uncertainty and are the same for the 50 ensemble members.

The state of the reservoir at each time step includes the following parameters:

- Logarithm of permeability values at pilot point locations (static variables): $\ln(k_{pi})$, $i = 1,...,5$
- Pressure in all grid cells (dynamic variables): P_j, $j = 1,...,2500$
- Water saturation in all grid cells (dynamic variables): Sw_j, $j = 1,...,2500$
- Bottom hole pressure at injection and production wells (measurements): BHP_k, $k = 1,...,5$
- Surface oil rate at production wells (measurements): SOR_k, $k = 1,...,4$
- Water cut at production wells (measurements): WCT_k, $k = 1,...,4$

Therefore, the state vector for the m^{th} ensemble member, ψ_m, is:

$$\psi_m = [\ln(k_{p1)},...,\ln(k_{p5}),P_1,...,P_{2500},Sw_1,...,Sw_{2500}, \quad BHP_1,...,BHP_5,SOR_1,...,SOR_4,WCT_1,...,WCT_4] \tag{6.2.}$$

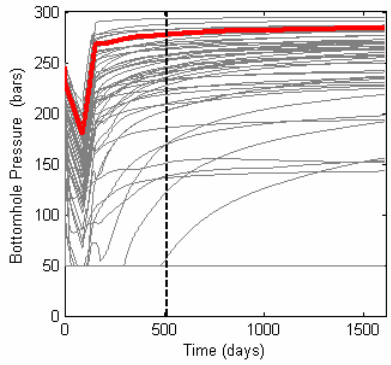
The history-matching period consists of 17 months (0-516 days): the target production data, displayed in Figure 3-2, are assimilated at the end of each month. To assess the capability of the updated models in predicting production after this period, we also consider a forecasting period lasting from day 517 to day 1612.

6.3.2 History-matching results

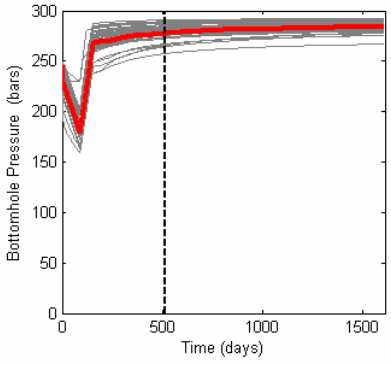
In this section, we present the results of the history matching process, applying EnKF to pilot point values.

The background ensemble for permeability realizations is the same as the initial ensemble used in Chapter 3 for the base case. In Figure 6-4, Figure 6-5 and Figure 6-6, we compare the performance of the initial and final ensembles about the four production wells (bottom hole pressure, surface oil rate and water cut, respectively). The first column displays the production responses simulated with the initial ensemble. The values simulated with the final ensemble over the whole time span are displayed in the second row. The history matching and forecast periods are separated by the dashed vertical line.

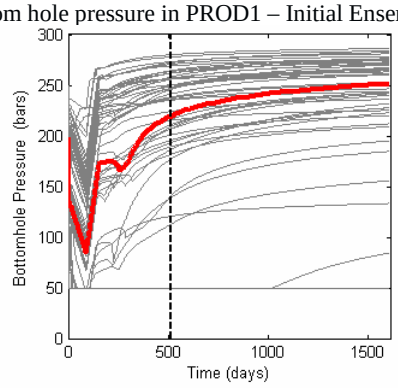
As can be seen in Figure 6-4 (first column), initial ensemble members have very different bottom hole pressure responses and some strongly differ from the reference data (red curve). The dispersion among the final ensemble members (second column) is lower and in general, the ensemble members can keep track of the reference model better than the initial ensemble both during the matching and forecasting periods. However, keeping in mind that measurement uncertainty is set to 5% of the reference bottom hole pressure at each assimilation step, some ensemble members do not provide an acceptable match at some wells. In particular, the final results for wells PROD2 and PROD4 are rather poor, considering the uncertainty level. We could observe better results, especially for wells PROD2 and PROD4, compared to Figure 6-4 if we restart the whole assimilation process with another ensemble. Again, this stresses the crucial role of the initial ensemble. In addition, the number and locations of pilot points play a major role. This will be further studied in the subsequent sections.



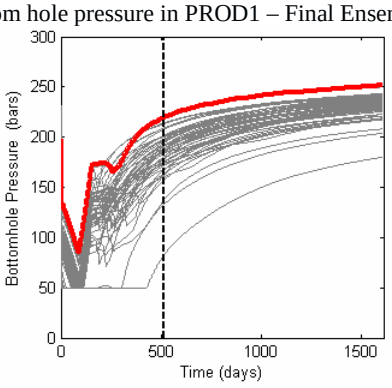
(a) Bottom hole pressure in PROD1 – Initial Ensemble



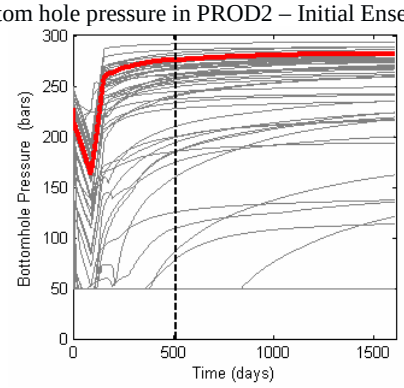
(e) Bottom hole pressure in PROD1 – Final Ensemble



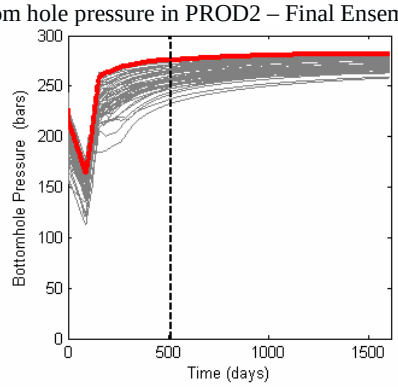
(b) Bottom hole pressure in PROD2 – Initial Ensemble



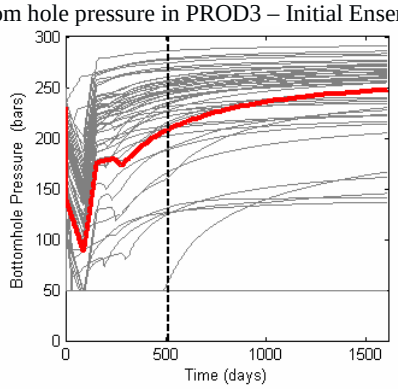
(f) Bottom hole pressure in PROD2 – Final Ensemble



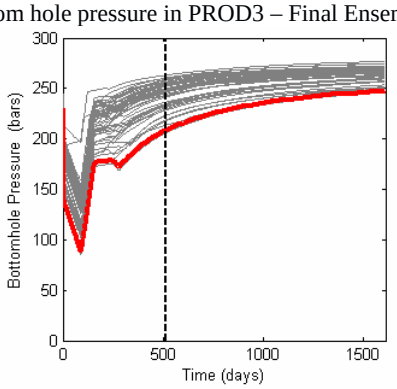
(c) Bottom hole pressure in PROD3 – Initial Ensemble



(g) Bottom hole pressure in PROD3 – Final Ensemble

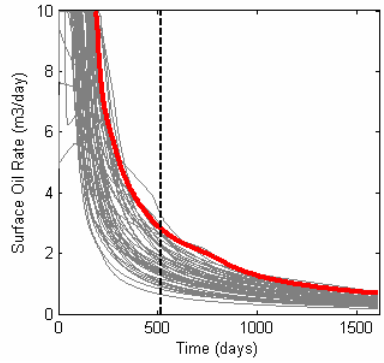


(d) Bottom hole pressure in PROD4 – Initial Ensemble

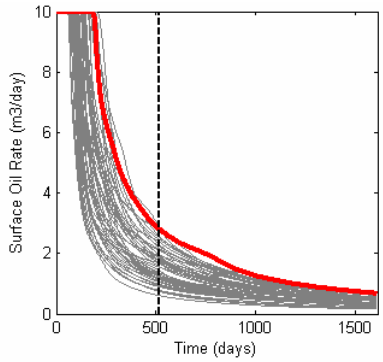


(h) Bottom hole pressure in PROD4 – Final Ensemble

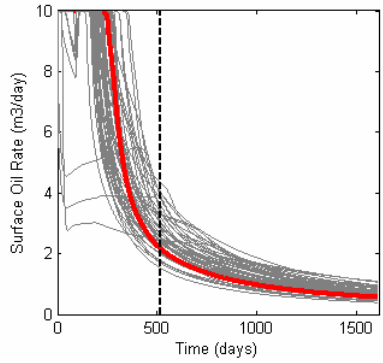
Figure 6-4 Bottom hole pressure for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. The reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



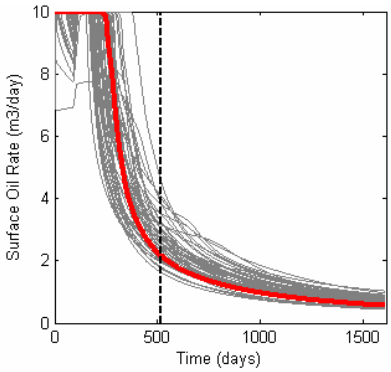
(a) Surface oil rate in PROD1 – Initial Ensemble



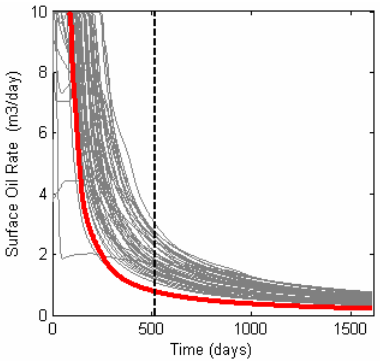
(e) Surface oil rate in PROD1 – Final Ensemble



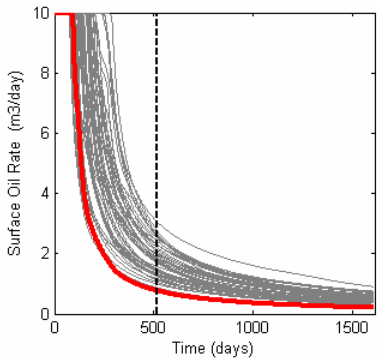
(b) Surface oil rate in PROD2 – Initial Ensemble



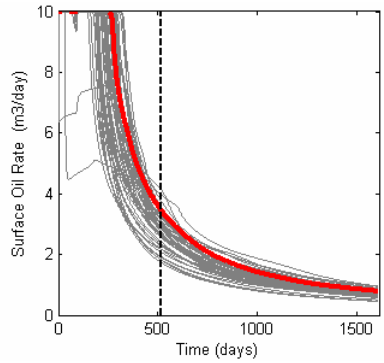
(f) Surface oil rate in PROD2 – Final Ensemble



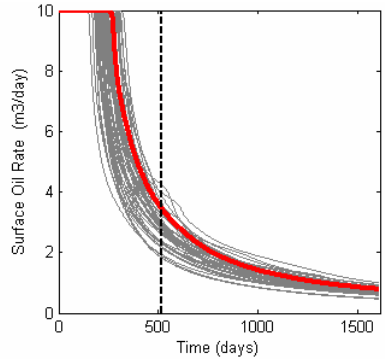
(c) Surface oil rate in PROD3 – Initial Ensemble



(g) Surface oil rate in PROD3 – Final Ensemble

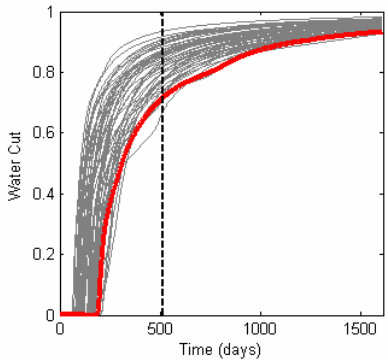


(d) Surface oil rate in PROD4 – Initial Ensemble

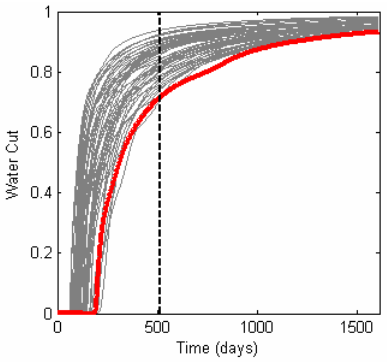


(h) Surface oil rate in PROD4 – Final Ensemble

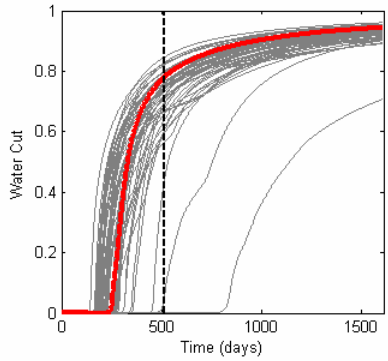
Figure 6-5 Surface oil rate for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. The reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



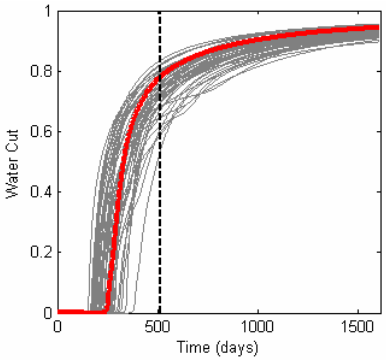
(a) Water cut in PROD1 – Initial Ensemble



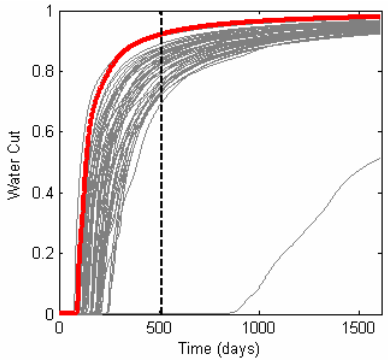
(e) Water cut in PROD1 – Final Ensemble



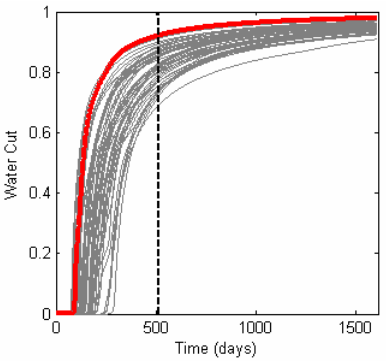
(b) Water cut in PROD2 – Initial Ensemble



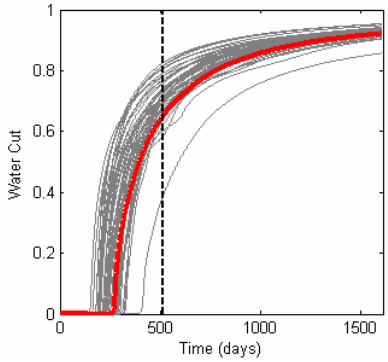
(f) Water cut in PROD2 – Final Ensemble



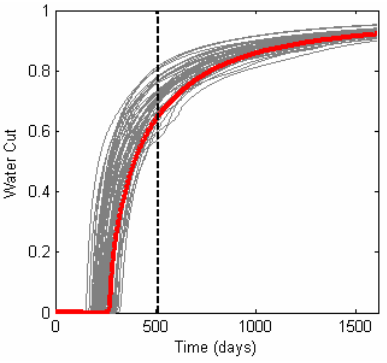
(c) Water cut in PROD3 – Initial Ensemble



(g) Water cut in PROD3 – Final Ensemble



(d) Water cut in PROD4 – Initial Ensemble

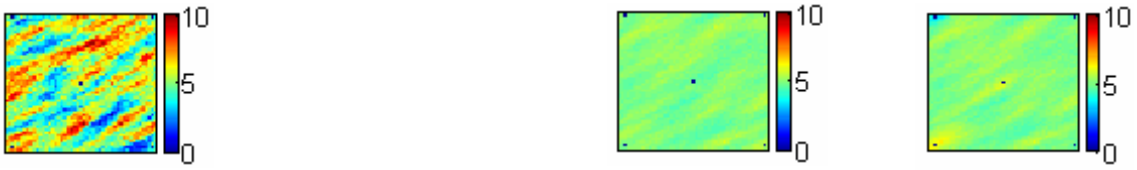


(h) Water cut in PROD4 – Final Ensemble

Figure 6-6 Water cut for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. The reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).

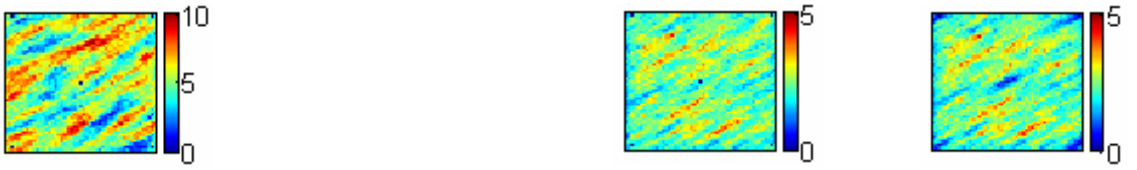
The matching parameter is permeability and it has an important influence on the reservoir pressure. Therefore, we expect better results for bottom hole pressure than for surface oil rate or water cut. This can be observed in Figure 6-5 and Figure 6-6: most of the outliers, *i.e.* results largely different from the reference results, are removed from the ensemble, but the dispersion is unchanged. Using another initial ensemble can slightly improve the results. We can also change the number and the locations of the pilot points, which would impact the results as will be shown further in this chapter. In terms of production forecast, the updated models are successful in keeping the track of the reference production data in the forecasting period, 517-1612 days. In addition, the spread of the final ensemble members around the reference model is slightly less compared with the initial ensemble members, especially for bottom hole pressure.

The assimilations and updates of the pilot point values by the EnKF make it possible to change flow ensemble members (permeability realizations) so that they start to reproduce some of the features of the reference permeability model (Figure 6-7). Among them are the high permeability region in the left bottom corner (around PROD1), the middle region (around INJ1) and the low permeability region (around PROD2) in the top left corner. However, the changes are not as obvious as those observed using the basic EnKF method in Chapter 3. This result was expected as by considering five pilot points, we drastically removed the number of freedom degrees (5 instead of 50×50).



(a) Reference model (b) Initial Ensemble (c) Final Ensemble

Figure 6-7 Reference permeability model and mean permeability computed from the initial and final ensembles. Pilot points are represented by dark plots.

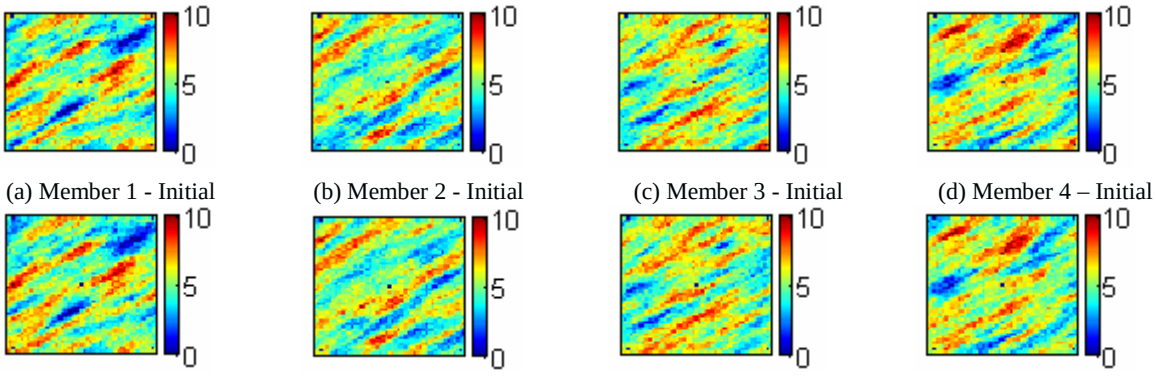


(a) Reference model (b) Initial Ensemble (c) Final Ensemble

Figure 6-8 Reference permeability model and variance of log permeability computed from the initial and final ensembles. Pilot points are represented by dark plots.

The variance within the updated flow ensemble during the assimilation process is shown in Figure 6-8. We can observe a strong decrease around pilot points. However, the variance does not vary in the rest of the reservoir. This clearly indicates that there are not enough pilot

points. In addition, permeability being usually known at wells, pilot points can be as well located elsewhere. This is the subject of section 6.4.3.



(a) Member 1 - Initial (b) Member 2 - Initial (c) Member 3 - Initial (d) Member 4 – Initial

(e) Member 1 - Final (f) Member 2 - Final (g) Member 3 - Final (h) Member 4 – Final

Figure 6-9 Four initial and final ensemble log permeability models. Pilot points are plotted in black.

Four realizations of the initial ensemble (without any assimilation) are shown in the first row of Figure 6-9, while their updated values (after performing 17 assimilations) are displayed in the second row. The following discussion is about the first member (Figure 6-9 (a) and (e)), but it could be extended to the three others. In the initial realization, well PROD1 is located in a low permeability region and wells PROD2, PROD3 and PROD4 in high permeability areas. However, the opposite can be observed in the reference model (Figure 6-7 (a)). The assimilation process contributes to adjust these permeability values so that they tend towards the reference values. Permeability in the other regions remains unchanged: more pilot points are required to improve the resemblance with the reference model.

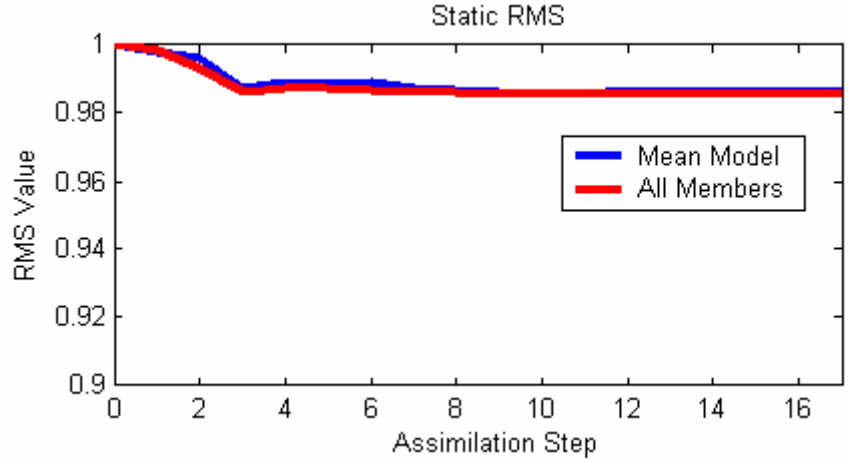


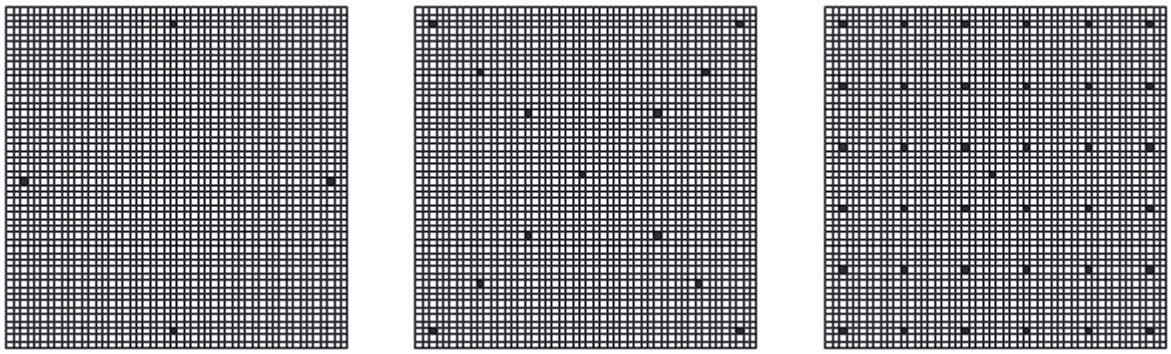
Figure 6-10 Static RMS: evolution of the logarithm of the permeability model with assimilation steps. The red curve corresponds to $RMS_{stat,1}$ and the blue one to $RMS_{stat,2}$.

The performance of the algorithm can be quantified in terms of RMS errors, defined for static data and production data on the basis of equations 3.2 through 3.5. For production data, the $RMS_{prod,1}$ computed for the final ensemble decreases to 89%, 87% and 87% of the initial

RMS value for the bottom hole pressure, surface oil rate and water cut, respectively. Figure 6-10 shows the evolution of the static RMS throughout the assimilations, considering both $RMS_{stat,1}$ and $RMS_{stat,2}$ (equations 3.4 and 3.5). The two curves are hardly distinguishable and show a decreasing trend. A larger decrease can be obtained by changing the number and position of the pilot points, as will be shown in the following section.

6.3.3 Effect of Number and Position of Pilot Points

As observed in the previous section, the ensemble members are adjusted around pilot points. In this section, three other configurations are investigated (Figure 6-11). Two of them were built adding more points to the 5-points configuration considered previously. The initial permeability ensemble is the same as in the previous section.



(a) Configuration 1 (b) Configuration 2 (c) Configuration 3

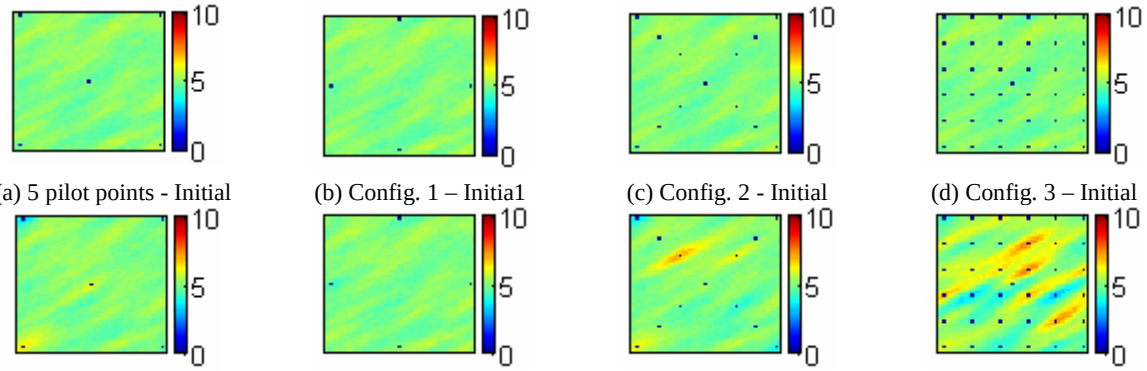
Figure 6-11 Three configurations for pilot point number and location. Pilot points are represented with the black dots.

Configuration 1 consists of 4 pilot points, placed at the middle of each side of the reservoir grid. The idea behind this setting is to investigate the capability of the algorithm in determining the permeability values around well locations if pilot points are located far from the wells. The pilot points are thus located in equidistance locations from the wells. The minimum distance between pilot points is 312m, which is larger than the smallest correlation length. However, in this configuration, the pilot points are at locations where the fluid flow is less significant.

In Configurations 2 and 3, more points are added to the initial 5-points configuration. 13 pilot points are considered in configuration 2, located at wells and at almost even intervals in the main flow direction between the injector and the producers. With this setting, we try to capture heterogeneities mainly in the flow direction. The minimum distance between two pilot points is 93m, but the anisotropy direction in the permeability realization is almost

perpendicular to the line joining the specified pilot points. Also, the correlation length in this direction is 40m. Thus, this pilot point setting should not suffer too much from the limitation relative to the minimal required lag between two pilot points. In configuration 3, 32 pilot points include the injector location plus a regular gridding including the producers. Almost all parts of the reservoir are thus covered. This setting violates, to some extent, the condition of at least one correlation length distance between pilot points. It was however tested here to see if small scale heterogeneities can be captured.

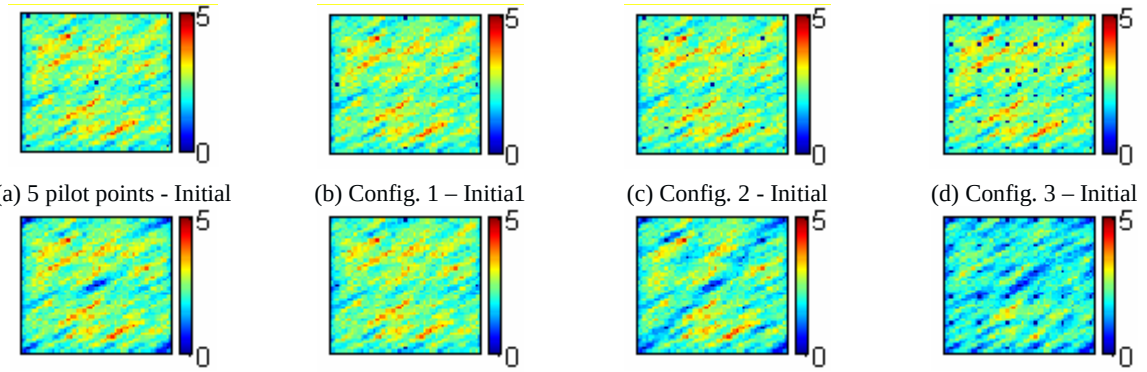
Figure 6-12 and Figure 6-13 represent the mean and variance obtained for the initial and final ensembles for the four configurations of interest.



(a) 5 pilot points - Initial (b) Config. 1 – Initia1 (c) Config. 2 - Initial (d) Config. 3 – Initial

(e) 5 pilot points - Final (f) Config. 1 – Fina1 (g) Config. 2 - Final (h) Config. 3 – Final

Figure 6-12 Mean model for logarithm of permeability distribution: a comparison between the 5 pilot point model configuration and Configurations 1 to 3.



(a) 5 pilot points - Initial (b) Config. 1 – Initia1 (c) Config. 2 - Initial (d) Config. 3 – Initial

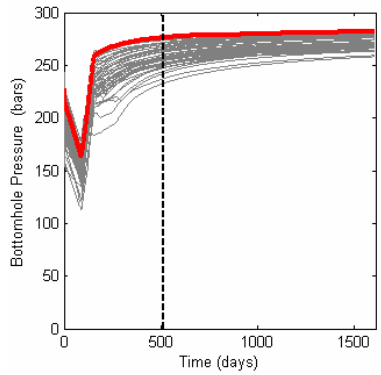
(e) 5 pilot points - Final (f) Config. 1 – Fina1 (g) Config. 2 - Final (h) Config. 3 – Final

Figure 6-13 Variance model for logarithm of permeability distribution: a comparison between the 5 pilot point model configuration and Configurations 1 to 3.

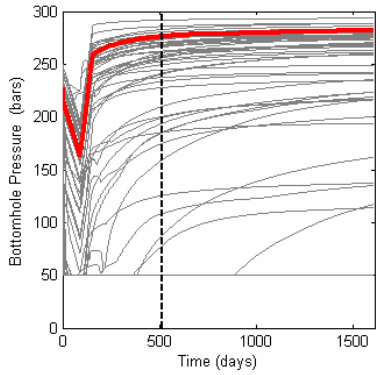
Considering pilot points located far from the wells (configuration 1), the algorithm does not succeed in regenerating the reference permeability field. The mean and variance remain almost unchanged (Figure 6-12 and Figure 6-13) and the spread of the simulated production response is not decreased. These results show that pilot points should be placed at locations where the fluid flow is significant.

Considering configurations 2 and 3, we can observe that some heterogeneity features appear around pilot points (Figure 6-12) and variance is reduced around these points (Figure 6-13). The same evolution as the one noticed for the 5-points configuration can be observed around wells. Since more points are considered in configuration 3, more heterogeneity features of the reference permeability model are captured in the final ensemble mean model. Also, the variance decreases all over the reservoir.

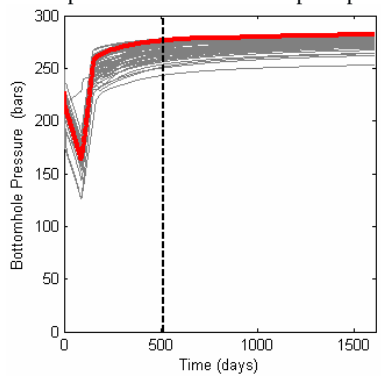
The match obtained for bottom hole pressure, surface oil rate and water cut (Figure 6-14 to Figure 6-16) is getting better when adding pilot points. The spread of the simulated production response within the final ensemble is progressively decreased from the 5-points configuration to configuration 3, leading to a successful match for configuration 3 with respect to the measurement error.



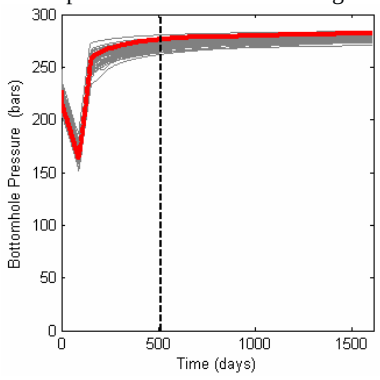
(a) Bottom hole pressure in PROD3 – 5 pilot points



(b) Bottom hole pressure in PROD3 – Configuration 1

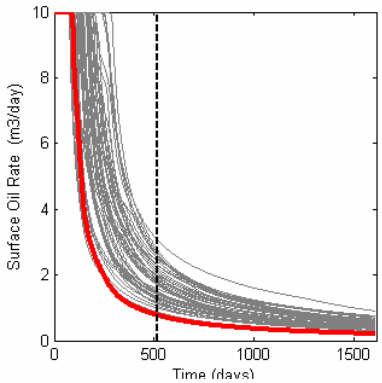


(c) Bottom hole pressure in PROD3 – Configuration 2

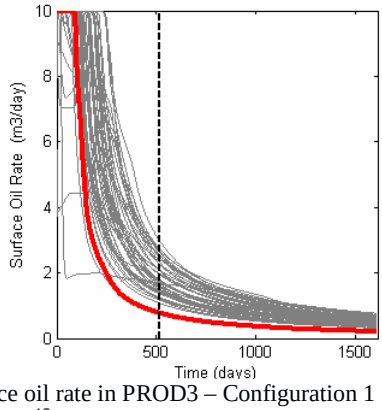


(d) Bottom hole pressure in PROD3 – Configuration 3

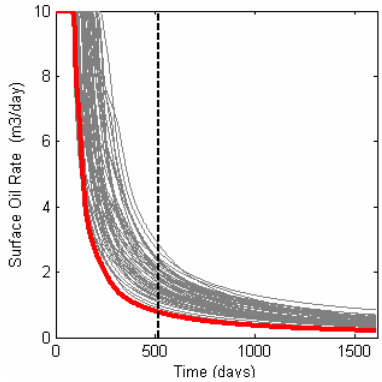
Figure 6-14 Bottom hole pressure in the final ensemble for well PROD3: comparison between (a) 5 pilot point configuration, (b) configuration 1, (c) configuration 2 and (d) configuration 3.



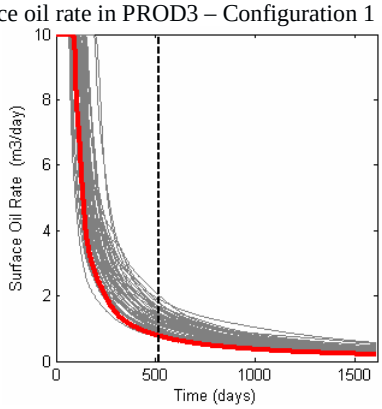
(a) Surface oil rate in PROD3 – 5 pilot points



(b) Surface oil rate in PROD3 – Configuration 1

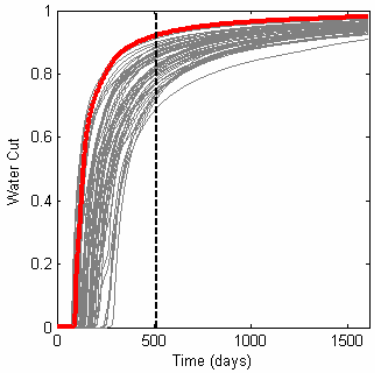


(c) Surface oil rate in PROD3 – Configuration 2

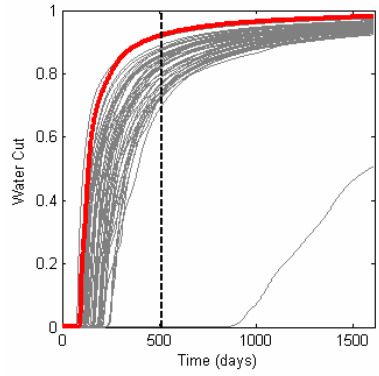


(d) Surface oil rate in PROD3 – Configuration 3

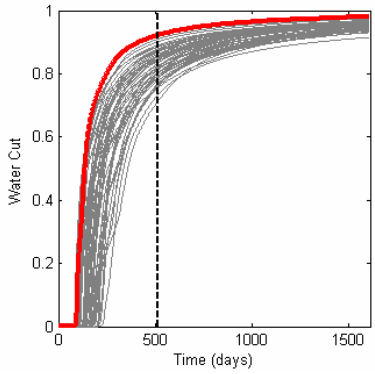
Figure 6-15 Surface oil rate in the final ensemble for well PROD3: comparison between (a) 5 pilot point configuration, (b) configuration 1, (c) configuration 2 and (d) configuration 3.



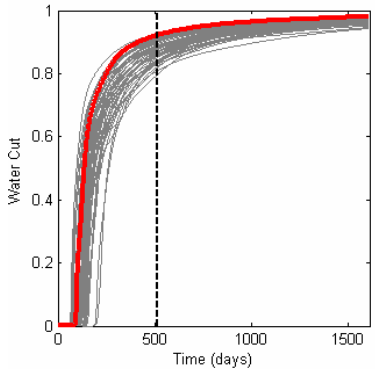
(a) Water cut in PROD3 – 5 pilot points



(b) Water cut in PROD3 – Configuration 1



(c) Water cut in PROD3 – Configuration 2



(d) Water cut in PROD3 – Configuration 3

Figure 6-16 Water cut in the final ensemble for well PROD3: comparison between (a) 5 pilot point configuration, (b) configuration 1, (c) Configuration 2 and (d) Configuration 3.

As already observed for the 5-points configuration, the static RMS curves computed over the mean model or all ensemble members have the same trend for configurations 1 to 3. Hence, we only consider the values computed over all ensemble members, $RMS_{stat,1}$ (equation 3.4), plotted in Figure 6-17.

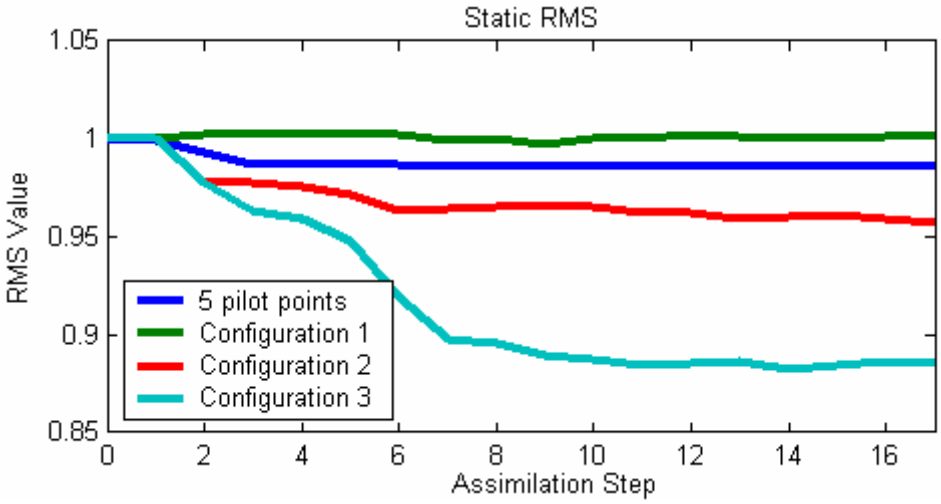


Figure 6-17 Static $RMS_{stat,1}$ comparison for different pilot point configurations.

As expected, static RMS is not reduced with configuration 1. Adding points, which basically cover the whole field, allows approximating better the reference permeability field.

The ratios of the final to initial $RMS_{prod,1}$ (equation 3.2) are given in Table 6-1. As previously, the worst results are obtained with configuration 1. They improve gradually when adding points.

Table 6-1 $RMS_{prod,1}$ for different pilot point settings – Comparison

Dynamic RMS	BHP	OPR	WCT
5 Pilot Points	0.89	0.88	0.87
Configuration 1	0.99	0.98	0.98
Configuration 2	0.62	0.88	0.89
Configuration 3	0.64	0.82	0.83

6.4 Conclusions

In this chapter, the theoretical framework for the pilot point parameterization method was discussed. An algorithm was proposed to combine the pilot point method with the EnKF history-matching method: the main idea is to consider only the static data at pilot points' locations in the state matrix and to update the whole model by kriging. Thus, the spatial variability of the property is preserved during assimilations. This algorithm was applied to a

simple 2D case study, highlighting the need to consider pilot points in regions impacted by flow to get satisfactory results. Also, adding more points contributes to improve the match and to better capture the reference model. However, this number is limited as if there are too many points, we can not ensure that the distance between two pilot points is more than the correlation length.

7 Gradual Deformation Method (GDM)

The gradual deformation method was first introduced by Hu (2000). Considering two Gaussian random functions with the same statistical properties, Hu proposed an adequate weighted combination of these functions which ensures that the resulting Gaussian random function satisfies the same properties. Using this technique, any realization of a Gaussian random function can be perturbed "smoothly" from a small number of parameters. Furthermore, the perturbed realization is still a realization of the random function.

In this chapter, we propose a method that combines the EnKF with the gradual deformation method, in order to preserve the spatial variability of the reservoir models during the assimilation process. The mathematical formulation of the GDM is recapped in section 7.1. The combination algorithm is described in section 7.2. It was applied to the 2D synthetic case introduced in Chapter 3. The results are given in section 7.4.

7.1 Mathematical formulation of the GDM

The GDM is based upon the weighted linear combination of Gaussian random functions with identical means and covariances. The weights are chosen properly so that the resulting Gaussian random function shares the same statistical properties as the other functions.

First, let us consider realizations y_1 and y_2 of two independent Gaussian random functions $Y1$ and $Y2$ with mean m and covariance C. Then, for any parameter t, y defined by:

$$y - m = (y_1 - m)\cos(\pi t) + (y_2 - m)\sin(\pi t) \qquad 7.1.$$

is the realization of a Gaussian random function Y with mean m and covariance C. t is the gradual deformation parameter. This formula is periodic on t, with a range of -1 to $+1$. For $t = 0$, y is the same as y_1 ; for $t = 1/2$, y is the same as y_2. Continuous perturbations of t thus result in continuous changes in realization y.

Formulation (7.1) can be extended to any number of realizations. Let us consider M independent Gaussian random functions ($Y1,...,YM$) with mean m and covariance C and a set of realizations $\{y_1, y_2,..., y_M\}$ of these functions. Then for any set of deformation parameters $t_1,...,t_{M-1}$,

$$y - m = \prod_{i=1}^{M-1} (y_1 - m)\cos(t_i) + \sum_{i=1}^{M-1} \sin(t_i) \prod_{j=i+1}^{M-1} \cos(t_j)(y_{i+1} - m) \qquad 7.2.$$

defines a realization of a Gaussian random function Y with mean m and covariance C.

The gradual deformation can be applied to Gaussian white noises. In this case, formula (7.2) is applied to the normal deviates of the Gaussian white noises. This can be used in particular to combine gradual deformation to Gaussian and non-Gaussian simulation algorithms such as the FFT-MA algorithm proposed by Le Ravalec *et al.* (2000), sequential simulations (Hu *et al.*, 2001) or object-based Boolean simulations (Hu, 2003). In the two last cases, deformation is applied to the related normal deviates involved in the simulation process. When using the FFT-MA algorithm, mean, variance, variogram model, *etc.*, can be perturbed independently from the underlying Gaussian white noises. Therefore, the gradual deformation can be applied either to the simulated realization itself or to the underlying Gaussian white noises. A major advantage when applying the GDM to normal deviates is to make local deformation possible.

When applying the same value of the deformation parameters to all grid blocks, the deformation is called "global". However, we can also assign different deformation parameter values to non overlapping regions of the reservoir. In this case, we talk about "local" gradual deformations (Hu, 2000). More details are provided further below.

For history-matching purposes, gradual deformation parameters can be considered as inversion parameters in an optimization process. However, considering the weighted sum of a finite number of realizations, only a part of the realization space can be investigated through the perturbation of the parameters. As a result, it may be of interest to consider several sets of realizations during the history-matching process, thus performing several optimizations. First, the deformation parameter values that minimize the objective function for a given set of realizations are determined. Then, during a second optimization, the previous best realization is combined to new realizations generated randomly, and the parameters values that minimize the objective function are searched for. The optimization process is stopped when the objective function has decreased to an acceptable level or when it converges to a stable level. More details on this gradual deformation based optimization workflow can be found in Le Ravalec-Dupin (2005).

Until now, it was only applied to variational methods. Many successful applications are reported in the literature. For instance, Le Gallo and Le Ravalec (2000) applied gradual deformation based optimizations to constrain 3D reservoir models to production data. They considered a synthetic case study and a real field case. The problem of constraining reservoir facies models to dynamic data is addressed in Zabalza-Mezghani *et al.* (2001), combining

gradual deformation to the truncated Gaussian simulation. Gradual deformation can also be used to the history-matching of fractured reservoirs. Finally, gradual deformation has been used for the joint history-matching of production and 4D seismic related data in Roggero *et al.* (2007).

7.2 Algorithm combining the EnKF and the GDM

In the present work, we propose an algorithm that combines gradual deformation with the EnKF sequential history-matching method. The basic idea is to consider as static parameters a set of gradual deformation parameters used to perturb an underlying ensemble of realizations. Each set of parameters provides a reservoir model used to simulate fluid flow evolution and production response. As a consequence, departure from the prior information, *e.g.* the geological model, is penalized and the final updated ensemble members are ensured to respect prior information. More specifically, updated ensembles are searched for in the space of realizations deduced from the prior information. Deformation parameters are applied either to the underlying Gaussian white noises or on the properties' realizations themselves. Let us note that the algorithm can also be used for facies distributions simulated with the truncated or pluri-Gaussian method. In this case, gradual deformation applies to the standard Gaussian distributions truncated according to thresholds values.

The algorithm is described in Figure 7-2. First, let us consider that gradual deformation applies to petrophysical properties distributions. The case of Gaussian white noises will be treated later on.

Let us consider an initial ensemble, called "background ensemble" in the remaining of this chapter, containing M realizations $b_i, i = 1,...,M$ for each static property of interest (porosity, permeability). It should be noted that the notion of background ensemble in this algorithm differs slightly from the one used in Chapter 7.

Then, N sets of $M-1$ gradual deformation parameters $\{t_{j,1},...,t_{j,M-1}\}, j = 1,...,N$ are generated randomly following a uniform distribution. These parameters will be considered as static variables when going through the assimilation process. Since the EnKF method requires Gaussian distributions, gradual deformation parameters are then submitted to a normal score transform. If several properties are updated during the EnKF assimilations, N sets of deformation parameters should be generated for each property. Then, at each assimilation time, the method can be followed in two steps:

1. Forecast

- Apply gradual deformation parameters to the background ensemble, resulting in N realizations of each property of interest $e_j, j = 1,...,N$ defined by:

$$e_j = \prod_{i=1}^{M-1}(b_1)\cos(t_{j,i}) + \sum_{i=1}^{M-1}\sin(t_{j,i})\prod_{k=i+1}^{M-1}\cos(t_{j,k})(b_{i+1}) \qquad 7.3.$$

  Here, b_i are the normalized and centralized members of the background ensemble. They yield a normalized and centralized realization e_j that is rescaled to the proper mean and variance. The resulting ensemble is called the "EnKF" ensemble (see also Figure 7-1).

- Forecast the performance of static models $e_j, j = 1,...,N$ and dynamic variables from a fluid flow simulation over the time step of interest. It provides a set of forecasted dynamic data and production responses.

2. Update

- Form the forecasted state vector for each EnKF ensemble member. The state vector for the j[th] EnKF member is composed of the set of gradual deformation parameters for this ensemble member, $t_{j,i}, i = 1,...,M-1$, the forecasted dynamic data (pressure, saturation,…) and simulated production data resulting from the fluid flow simulation on e_j,
- Apply the EnKF update step to the forecasted state matrix
- The resulting set of updated deformation parameters and dynamic data are used as initial values for the forecast at the following time step.

Let us note that, at the first assimilation, the dynamic variables are given by the initial equilibrium in the reservoir models.

In the case where gradual deformation applies to Gaussian white noises, the background ensemble consists of M Gaussian white noises $b_i, i = 1,...,M$ used to generate the property of interest (using the FFT-MA algorithm for instance). The N sets of deformation parameters $\{t_{j,1},...,t_{j,M-1}\}, j = 1,...,N$ are used to combine these Gaussian white noises using formula 7.3. The N resulting Gaussian white noises are then transformed to N realizations $e_j, j = 1,...,N$ of the property of interest applying some geostatistical algorithm such as FFT-MA.

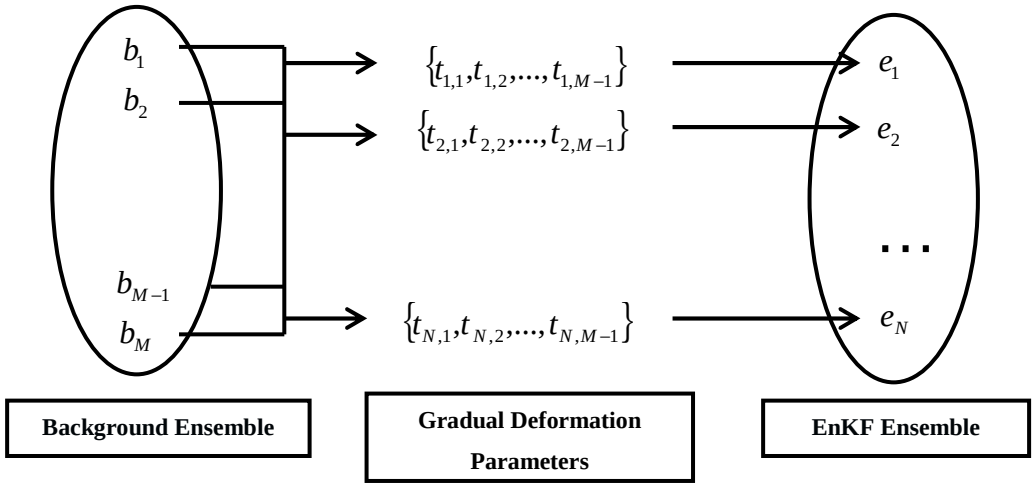


Figure 7-1 Algorithm combining the Ensemble Kalman Filter and the Gradual Deformation parameterization method.

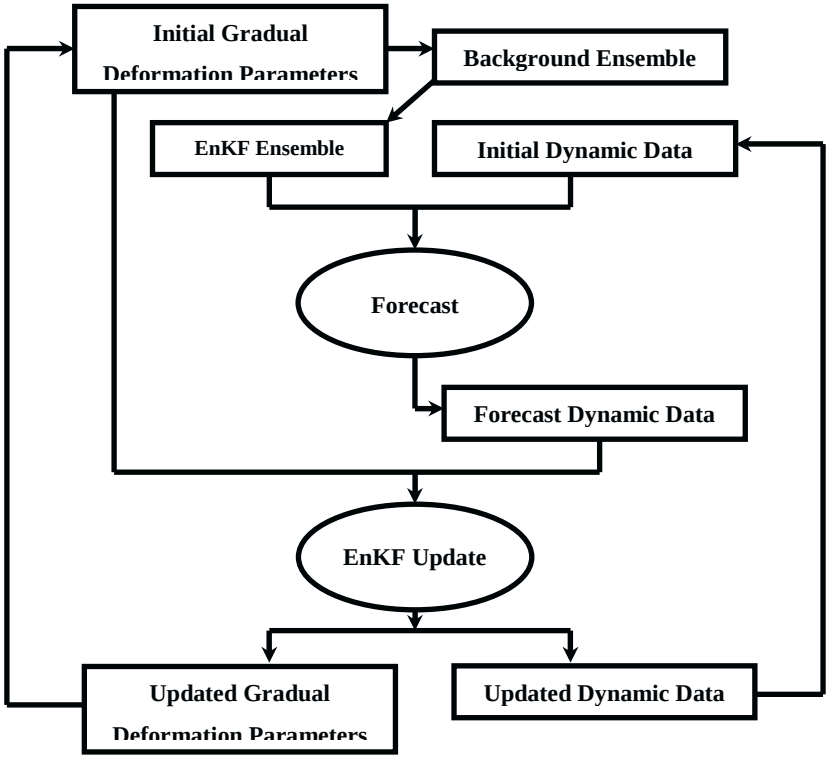


Figure 7-2 Workflow for the algorithm combining the EnKF with the gradual deformation parameterization method

In this algorithm, as the reservoir models are obtained applying gradual deformation to the background ensemble, it is ensured that they all respect the prior information and the initial spatial variability. As a result, some of the drawbacks already identified for the EnKF, such as the overestimation of petrophysical properties, should be avoided. In addition, the size of the background ensemble can be changed independently from the size of the EnKF ensemble. It

means that the search space for constrained reservoir models can be increased without adding any fluid flow simulation in the forecast stage of the EnKF. We may call this effect a "virtual" ensemble size increase. It will be investigated in section 7.3.1.2.

However, it should be mentioned that the algorithm also increases the degree of nonlinearity in the problem. The reason is the nonlinear dependence of the gradual deformation parameters $t_{j,i}, i = 1,...,M-1$ and combined realizations $b_i, i = 1,...,M$ on the final realization e_j due to the gradual deformation formulation. Moreover, the EnKF ensemble realizations may exhibit common features as they are generated from the same background ensemble. Therefore, the forecast error covariance matrix may not represent the uncertainties in the model to an adequate level.

In some cases, global gradual deformation does not provide enough flexibility to obtain a good match on the data. For instance, an improvement of the match at some wells can deteriorate a satisfactory match at another well. In this case, local deformations provide more flexibility and can make the history-matching process more efficient. The algorithm proposed above to combine gradual deformation and the EnKF can be easily extended to local deformations. In that case, gradual deformation should be applied to Gaussian white noises.

Assume that the reservoir model is divided into Z non overlapping regions. This zonation is performed either on a regular basis or based on some knowledge about the reservoir. Each zone consists of $Num_k, k = 1,...Z$ grid blocks such that:

$$\sum_{k=1}^{Z} Num_k = \text{ Total number of grid blocks} \qquad 7.4.$$

Consider that the size of the EnKF ensemble is set to N and the size of the "global" background ensemble is set to M. Then, N sets of $M-1$ gradual deformation parameters $\left\{t^k_{j,1},...,t^k_{j,M-1}t_{j,M-1}\right\}, j = 1,...,N$ are generated randomly for each region $k = 1,..,Z$ and used to combine the Gaussian white noises $b_i, i = 1,...,M$ of the background ensemble. It results in N Gaussian white noises $\tilde{b}_j, j = 1..N$ used to generate the N realizations $e_j, j = 1,...,N$. defined by:

$$\tilde{b}_j(x) = \prod_{i=1}^{M-1} b_1(x)\cos(t_{j,i}^{k(x)}) + \sum_{i=1}^{M-1}\sin(t_{j,i}^{k(x)})\prod_{l=i+1}^{M-1} b_{i+1}(x)\cos(t_{j,l}^{k(x)}) \qquad 7.5.$$

for all grid block x, $k(x) \in \{1,...,Z\}$ being the number of the region x belongs to. This variant of the algorithm will be investigated in section 7.3.1.3.

7.3 Two Dimensional Case Study

The 2D case study introduced in Chapter 3 is used to assess the performance of the algorithm combining the EnKF and the gradual deformation parameterization method. The aim of the assimilation process is to obtain log permeability realizations, which reproduce the well production data of the reference permeability model. The production and assimilation criteria, assimilation time interval and measurement uncertainty level are the same as for the base case. Gradual deformation is applied to the underlying Gaussian white noises in section 7.3.1 and to the static properties of interest in section 7.3.2. The effect of the size of the background ensemble is investigated in section 7.3.1.2 and the effect of the local deformations in section 7.3.1.3.

7.3.1 History-Matching with gradual deformation applied to the Gaussian White Noises

7.3.1.1 Base Case Study

The experiment in this section is considered as the base case for the remaining of the chapter. The size of the EnKF ensemble is set to $N = 50$. The background ensemble encompasses $M = 100$ Gaussian white noises. Therefore, a set of 99 gradual deformation parameters is associated to each EnKF ensemble member. FFT-MA is applied to the combined white noises to generate the set of N realizations for the property of interest, using the same characteristics as the reference permeability field. However, it is ensured that the reference $\ln k$ is not among these permeability realizations.

The EnKF state vector for each ensemble member, m, includes:

- The set of gradual deformation parameters (static variable): $t_{m,i}$, $i = 1,...,M-1$
- Pressure in all grid cells (dynamic variable): $P_{m,j}$, $j = 1,...,2500$
- Water saturation in all grid cells (dynamic variable): $Sw_{m,j}$, $j = 1,...,2500$
- Bottom hole pressure at injection and production wells: $BHP_{m,k}$, $k = 1,...,5$
- Surface oil rate at production wells: $SOR_{m,k}$, $k = 1,...,4$
- Water cut at production wells: $WCT_{m,k}$, $k = 1,...,4$

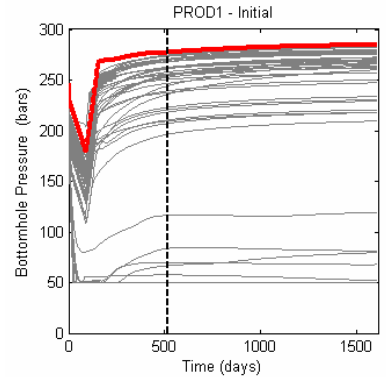
Therefore, the state vector for the m^{th} ensemble member, ψ_m, is:

$$\psi_m = [t_{m,1},...,t_{m,M-1}, P_{m,1},...,P_{m,2500}, Sw_{m,1},...,Sw_{m,2500}, \\ BHP_{m,1},...,BHP_{m,5}, SOR_{m,1},...,SOR_{m,4}, WCT_{m,1},...,WCT_{m,4}] \quad 7.6.$$

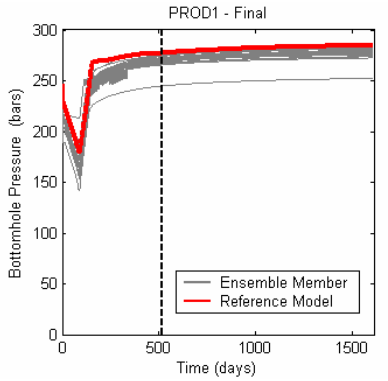
The production data are assimilated with the proposed algorithm at the end of each month during 17 months. The reference production data are as displayed in Figure 3-2. Moreover, the forecast capability of the updated ensemble members is further studied in the prediction period which extends between 517 to 1612 days.

We compare the performance of the initial and final EnKF ensembles in terms of matching the reference well production data and reproducing the reference $\ln k$ model.

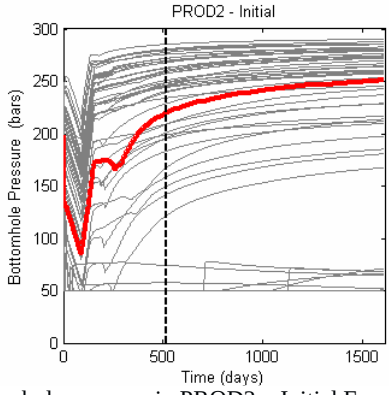
The well production responses obtained by the initial and final EnKF ensemble members are shown in Figure 7-3, Figure 7-4 and Figure 7-5 for bottom hole pressure, surface oil rate and water cut respectively. As can be observed, the outliers, *i.e.* production data far from the reference data, are removed by performing assimilation and the spread within the simulated responses is reduced. The match for bottom hole pressure in wells PROD1 and PROD4 is of acceptable quality. However, the match is not at all satisfactory for well PROD2. For well PROD3, it is at least better than for the initial ensemble. The results are poorer for the surface oil rate and water cut: the matches obtained for wells PROD3 and PROD4 only are satisfactory while the ones for PROD1 and PROD2 are rather poor. Globally, the quality of the match is not as good as the one obtained with the traditional EnKF method (Chapter 3). The use of other initial background ensembles and sets of initial gradual deformation parameters for the same background ensemble did not provide more satisfactory results.



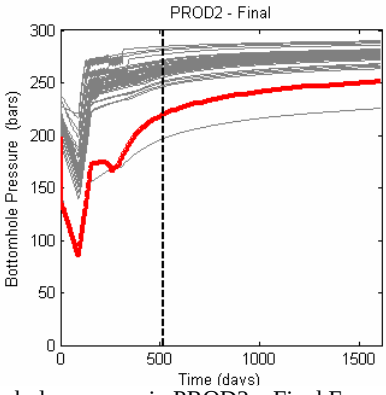
(a) Bottom hole pressure in PROD1 – Initial Ensemble



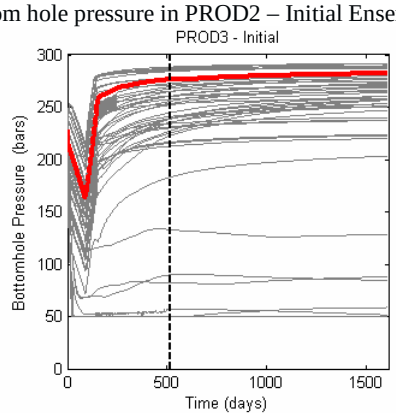
(e) Bottom hole pressure in PROD1 – Final Ensemble



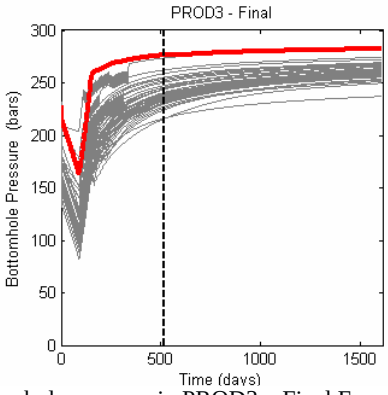
(b) Bottom hole pressure in PROD2 – Initial Ensemble



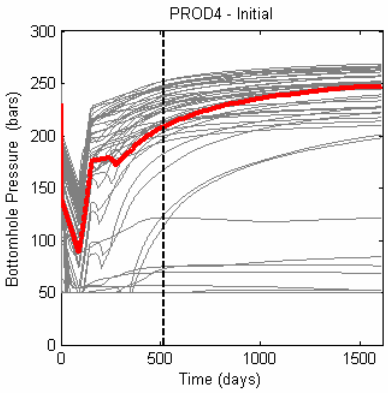
(f) Bottom hole pressure in PROD2 – Final Ensemble



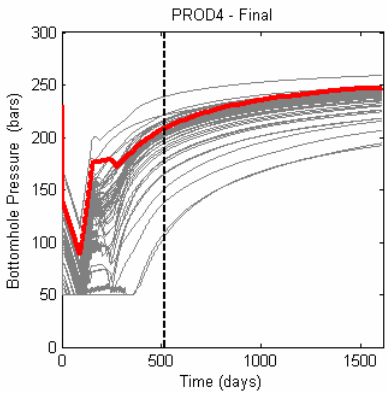
(c) Bottom hole pressure in PROD3 – Initial Ensemble



(g) Bottom hole pressure in PROD3 – Final Ensemble

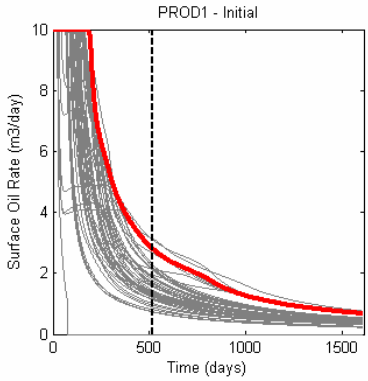


(d) Bottom hole pressure in PROD4 – Initial Ensemble

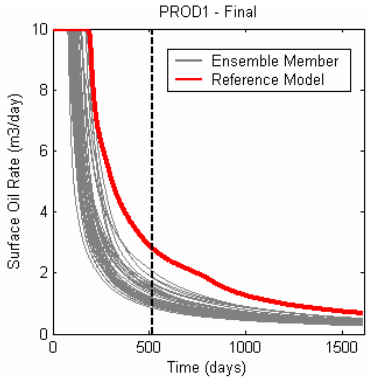


(h) Bottom hole pressure in PROD4 – Final Ensemble

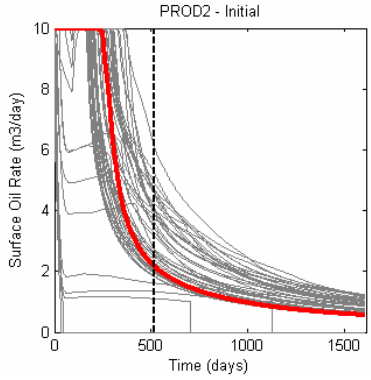
Figure 7-3 Bottom hole pressure for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



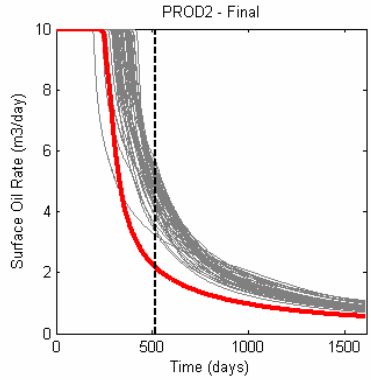
(a) Surface oil rate in PROD1 – Initial Ensemble



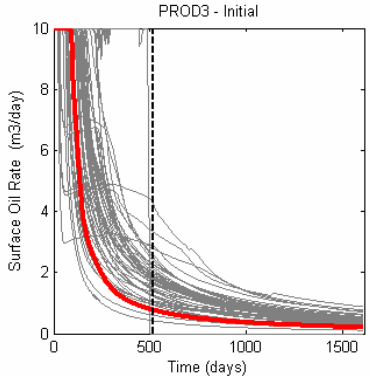
(e) Surface oil rate in PROD1 – Final Ensemble



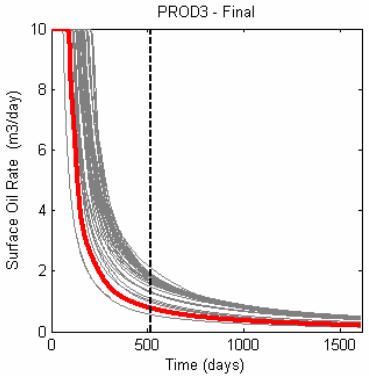
(b) Surface oil rate in PROD2 – Initial Ensemble



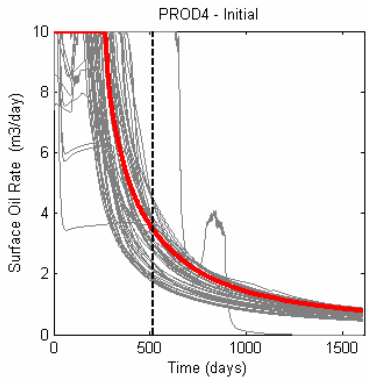
(f) Surface oil rate in PROD2 – Final Ensemble



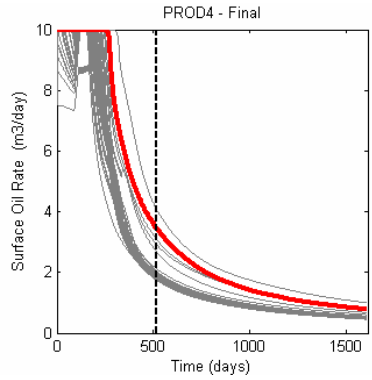
(c) Surface oil rate in PROD3 – Initial Ensemble



(g) Surface oil rate in PROD3 – Final Ensemble

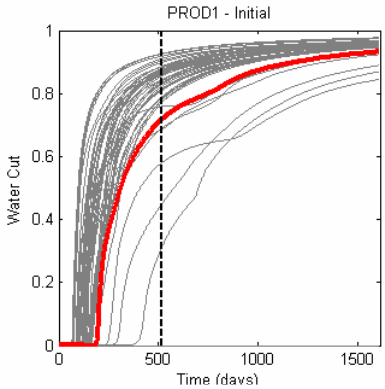


(d) Surface oil rate in PROD4 – Initial Ensemble

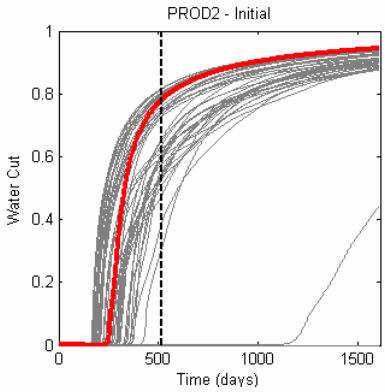


(h) Surface oil rate in PROD4 – Final Ensemble

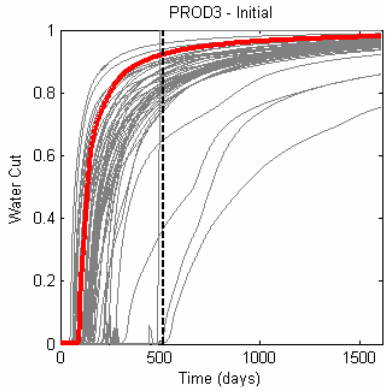
Figure 7-4 Surface oil rate for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



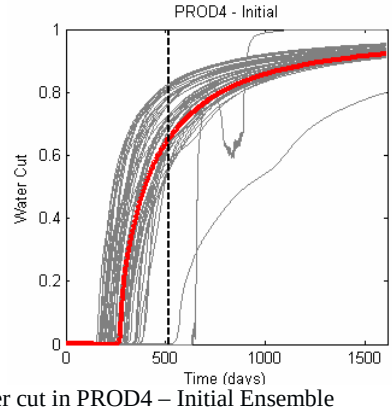
(a) Water cut in PROD1 – Initial Ensemble



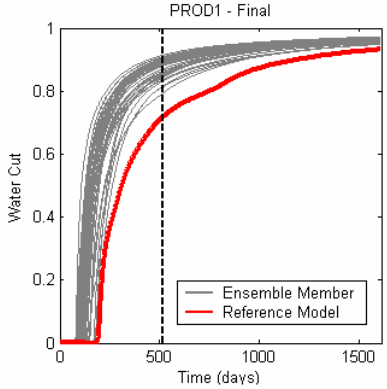
(b) Water cut in PROD2 – Initial Ensemble



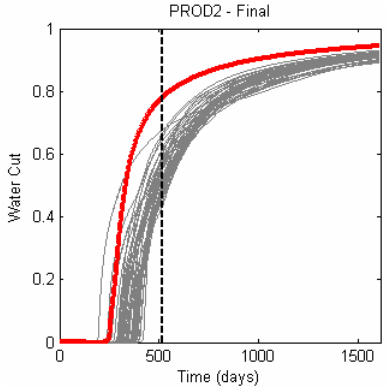
(c) Water cut in PROD3 – Initial Ensemble



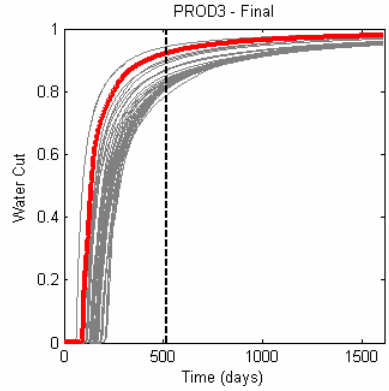
(d) Water cut in PROD4 – Initial Ensemble



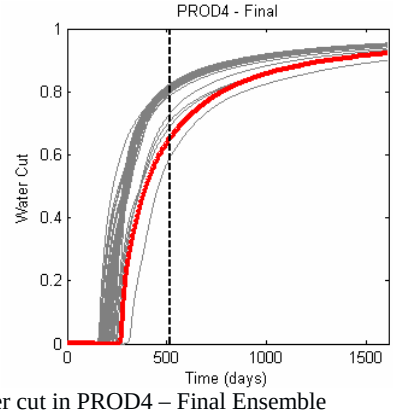
(e) Water cut in PROD1 – Final Ensemble



(f) Water cut in PROD2 – Final Ensemble



(g) Water cut in PROD3 – Final Ensemble

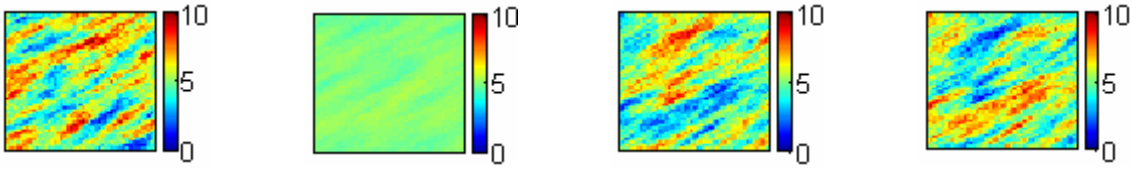


(h) Water cut in PROD4 – Final Ensemble

Figure 7-5 Water cut for the initial ensemble (a) PROD1, (b) PROD2, (c) PROD3, (d) PROD4 and the final ensemble (e) PROD1, (f) PROD2, (g) PROD3 and (h) PROD4. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).

This rather poor outcome may be due to inherent properties of the algorithm. However, it may be improved by introducing more flexibility, such as larger background ensembles and local deformations.

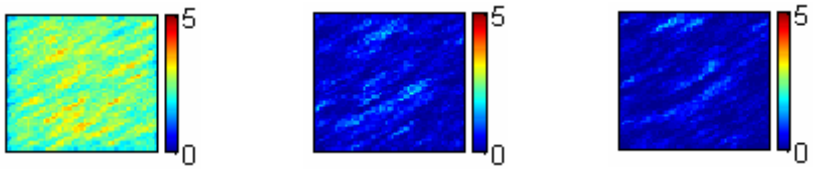
Two main issues can be highlighted for the algorithm. First, gradual deformation is not a linear parameterization: the relation between the deformation parameters and the final combined realizations is not linear. Therefore, the updated gradual deformation parameters after each assimilation might not correspond to the best combined realizations to be used for the following assimilation step. Figure 7-6 represents the mean $\ln k$ model obtained by taking an average over the EnKF permeability realizations of the initial ensemble, the ensemble after 12 assimilations and the ensemble after 17 assimilations. It can be observed that the corrections obtained on $\ln k$ after 12 assimilations are not preserved in the final model, implying once more that the nonlinear parameterization may be inappropriate for the assimilations with the EnKF.



(a) Reference Model (b) Initia1 Model (c) 12th Assimilation (d) 17th Assimilation

Figure 7-6 ln k evolution through time: (a) Reference model, (b) Initial mean model, (c) Mean model after 12th assimilation, (d) Mean model after 17th assimilation.

Finally the set of gradual deformation parameters may converge to certain values and result in a very low variance between the EnKF ensemble members (Figure 7-7). Therefore, the algorithm converges to a set of final updated models which are by far different from the reference model.



(a) Initia1 Model (b) 12th Assimilation (c) 17th Assimilation

Figure 7-7 Variance of ln k evolution through time: (a) Initial model (b) Variance model after 12th assimilation (d) Variance model after 17th assimilation.

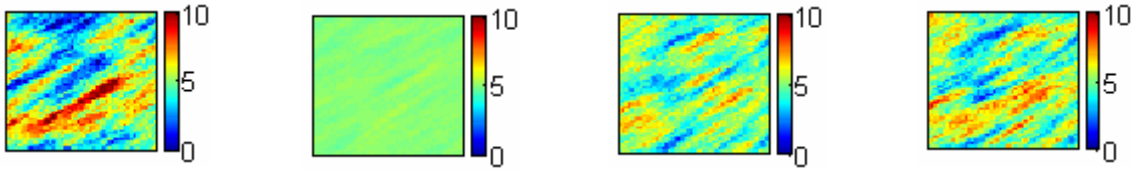
The second issue is the limited size of the search space for this optimization problem induced by the limited size of the background ensemble. Therefore, if the reference model is not kept within the possible combinations of the background ensemble members, the algorithm may not be capable of capturing the reference model. Moreover, the EnKF ensemble members are limited to realizations which can be defined through gradual deformation parameterization.

This makes the algorithm even more restrictive. This issue was further developed with the following experiment. A new reference permeability model was built by applying the FFT-MA algorithm to one of the background ensemble members. Then, applying the algorithm with the same assimilation scenario as before, the quality of the match was improved for all production data at all wells (Table 7-1).

Table 7-1 $RMS_{prod,1}$ for the initial and final EnKF ensembles obtained by the combined EnKF and GDM algorithm

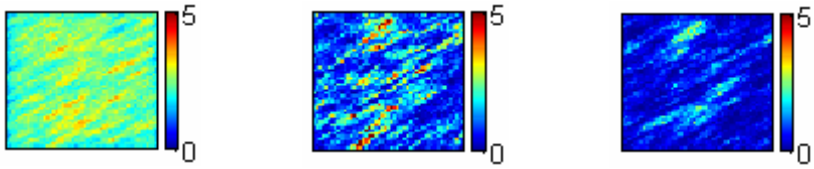
Dynamic RMS	BHP	SOR	WCT
Initial	6.8556	0.2794	0.0279
Final	5.6332	0.1873	0.0187
Final / Initial	0.8205	0.6704	0.0670

As can be observed in Figure 7-8, several features of the new reference permeability model are reproduced and the variance (Figure 7-9) has decreased by performing assimilation. In such a case, the proposed algorithm seems successful in converging toward the reference model. However, the traditional EnKF is still more successful.



(a) Reference Model (b) Initia1 Model (c) 12th Assimilation (d) 17th Assimilation

Figure 7-8 ln k evolution through time for the new reference model: (a) Reference model, (b) Initial mean model, (c) Mean model after 12th assimilation, (d) Mean model after 17th assimilation.



(a) Initial Model (b) 12th assimilation (c) 17th assimilation

Figure 7-9 Variance of ln k evolution through time for the new reference model: (a) Initial model (b) Variance model after 12th assimilation (d) Variance model after 17th assimilation.

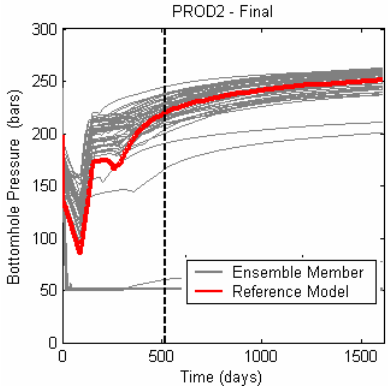
In the following subsections, we consider some general variants of the algorithm combining the EnKF with gradual deformation to improve its performance.

7.3.1.2 Effect of background ensemble size

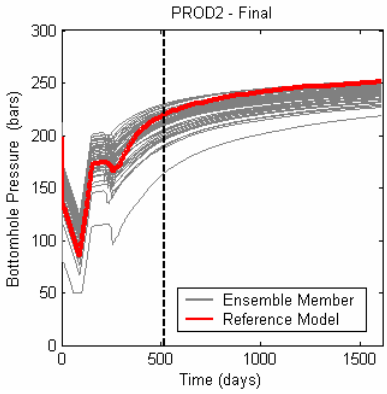
A key parameter in the algorithm is the size of the background ensemble. The bigger the size of the background ensemble, the larger the search space for updated models. In addition, as a general rule, the performance of parameterization methods improves as they include more

parameters. Here, for a background size of M, $M-1$ gradual deformation parameters are considered. As the GDM is regarded as a parameterization method, $M-1$ must be smaller than the total number of grid cells in the model.

We performed several experiments with various sizes of the background ensemble, considering successively $M=100$, 1000 and 2500. In all experiments, the assimilation scenario and the size N of the EnKF ensemble are the same as before. Bottom hole pressure, surface oil rate and water cut simulated with the final ensembles in well PROD2 are shown in Figure 7-10, Figure 7-11 and Figure 7-12 for $M=1000$ and 2500, and in Figure 7-3 (f), Figure 7-4 (f) and Figure 7-5 (f) for $M=100$.

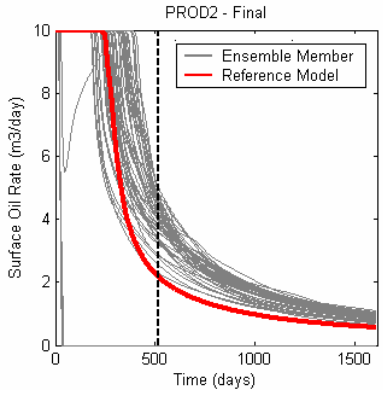


(a) Bottom hole pressure in PROD2 – Final Ensemble Background ensemble size: 1000

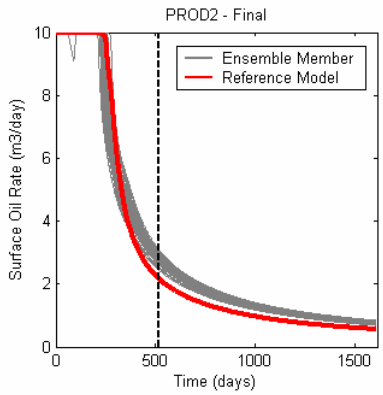


(b) Bottom hole pressure in PROD2 – Final Ensemble Background ensemble size: 2500

Figure 7-10 Bottom hole pressure for final ensembles (a) background ensemble size: 1000, (b) background ensemble size: 2500, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



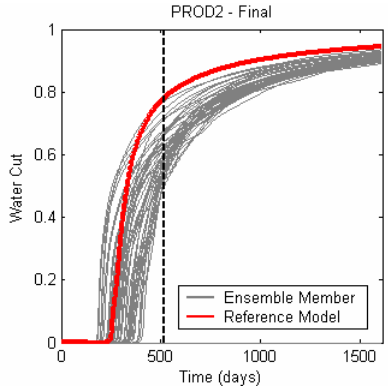
(a) Surface oil rate in PROD2 – Final Ensemble Background ensemble size: 1000



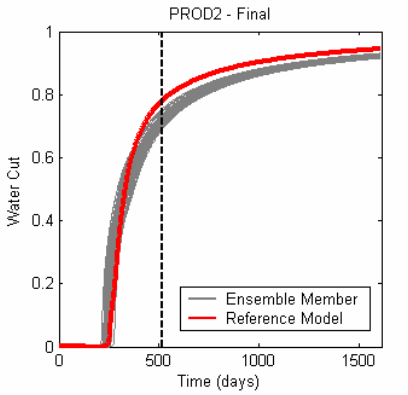
(b) Surface oil rate in PROD2 – Final Ensemble Background ensemble size: 2500

Figure 7-11 Surface oil rate for final ensembles (a) background ensemble size: 1000 , (b) background ensemble size: 2500, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).

As the background ensemble size increases, the assimilation results approximate better the reference production data. However, they do not outperform the traditional EnKF method.



(a) Water cut in PROD2 – Final Ensemble
Background ensemble size: 1000



(e) Water cut in PROD2 – Final Ensemble
Background ensemble size: 2500

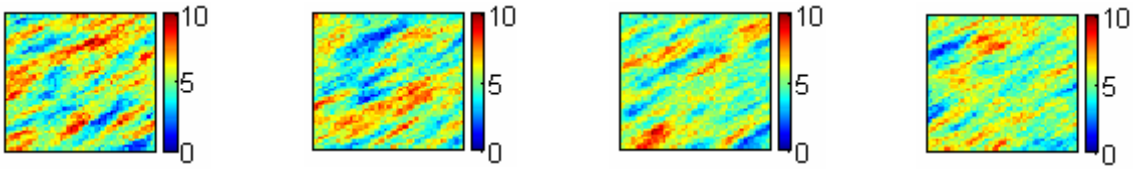
Figure 7-12 Water cut for final ensembles (a) background ensemble size: 1000, (b) background ensemble size: 2500, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red . The black dashed line separates the history-matching period (0-516 days) from forecasting period (517-1612 days).

The $RMS_{prod,1}$ for well production data (equation 3.2) is given in Table 7-2. In order to make comparison easier, the RMS values for the final ensembles are normalized to the final RMS value obtained with a background ensemble of size 100. The RMS value for all production data decreases significantly when increasing the size of the background ensemble. This is a straight consequence of the increase in search space. Thus we are capable of regenerating the trend of the reference well production data. However, it should be noted that, as for traditional EnKF, the choice of the background ensemble and initial gradual deformation parameters strongly impact the results.

Table 7-2 Effect of the background ensemble size: Value of the normalized final $RMS_{prod,1}$ obtained by increasing the size of the background ensemble. Normalization is performed with respect to the final values for the background ensemble of size 100.

Background Ensemble Size	BHP	SOR	WCT
100	1.0	1.0	1.0
1000	0.7390	0.7764	0.7494
2500	0.3593	0.4124	0.4134

Figure 7-13 represents the mean permeability of the final EnKF models for different background ensemble sizes.



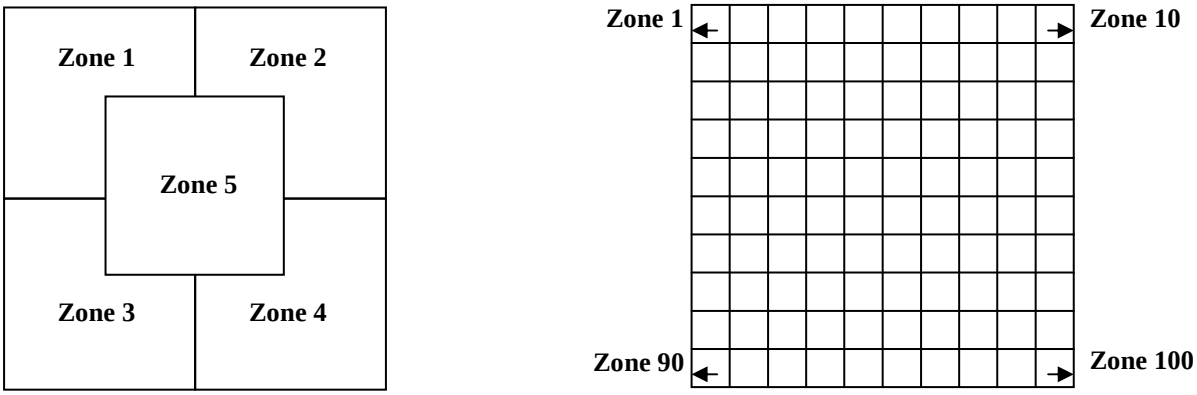
(a) Reference Model (b) 100 (c) 1000 (d) 2500

Figure 7-13 Final Mean model for the ln kh for different background ensemble sizes: (a) Reference model, (b) Background ensemble of size 100, (c) Background ensemble of size 1000, (d) Background ensemble of size 2500

Considering the background ensemble of size 100 as the base case, the $RMS_{stat,1}$ (equation 3.4) value was reduced to 90% and 85% of the base case value by increasing the background ensemble size to 1000 and 2500. However, none of them can be considered as a regeneration of the reference model. In addition, the regenerated common features are not the same for the three studied cases. This may be considered as the result of the nonlinear parameterization method used in this algorithm. Moreover, the initial EnKF ensemble is not the same in these cases. Overall, the results may be improved by trying different background ensembles.

7.3.1.3 Effect of local Gradual Deformation with the EnKF

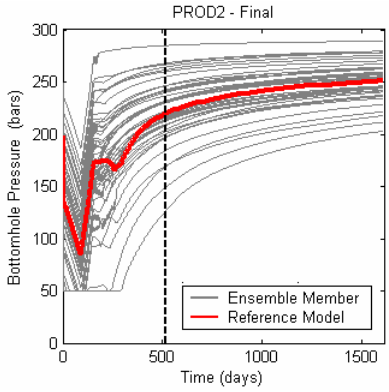
In some applications, it may be desirable to locally modify the value of the realizations on some areas of the reservoir. This provides more flexibility and should improve the data match. In this study, we consider two different configurations for regions' definitions on the 2D reservoir grid. In the first one (Figure 7-14(a)), the reservoir is split into 5 non regular zones surrounding wells. Zones 1 to 4 are made up of 481 grid blocks and zone 5 of 576 grid blocks. In the second configuration (Figure 7-14(b)), the grid is divided in 100 cubic zones, each one with 25 grid blocks.



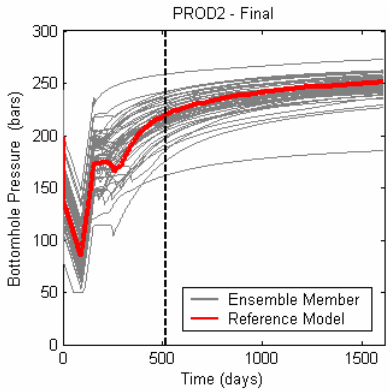
(a) 5 zones configuration (b) 100 zones configuration

Figure 7-14 Zone configurations for combining the local gradual deformation and the EnKF: (a) 5 zones configuration and (b) 100 zones configuration.

The assimilation scenario is the same as before. The global background ensemble has a size of 100 and the EnKF ensemble has a size of 50. The EnKF state vector is the same as in section 7.4.1, except that the set of static parameters now consists of the local gradual deformation parameters in each zone.

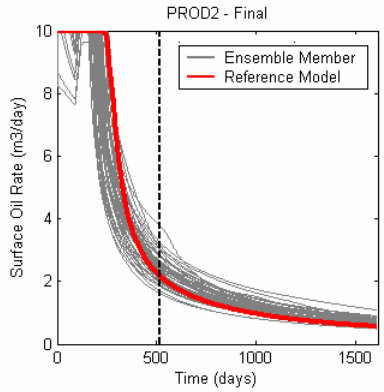


(a) Bottom hole pressure in PROD2 – Final Ensemble 5 zones configuration

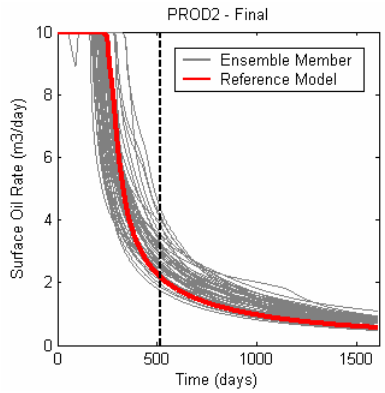


(b) Bottom hole pressure in PROD2 – Final Ensemble 100 zones configuration

Figure 7-15 Bottom hole pressure for final ensembles (a) 5 zones configuration , (b) 100 zones configuration, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).



(a) Surface oil rate in PROD2 – Final Ensemble 5 zone configuration

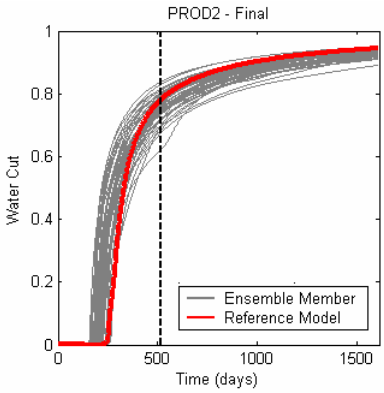


(b) Surface oil rate in PROD2 – Final Ensemble 100 zone configuration

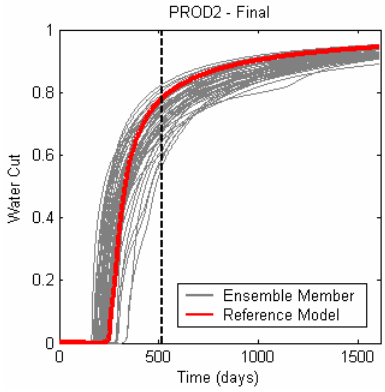
Figure 7-16 Surface oil rate for final ensembles (a) 5 zones configuration , (b) 100 zones configuration, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separated the history-matching period (0-516 days) from forecasting period (517-1612 days).

Figure 7-15, Figure 7-16 and Figure 7-17 represent the well production data for the final updated ensembles for the two zones configurations. The corresponding results obtained with only one zone and the same size of the background and EnKF ensembles are shown in Figure 7-3 (f), Figure 7-4 (f) and Figure 7-5 (f). It can be observed that the match for bottom hole pressure improves significantly when increasing the number of regions. In terms of surface oil

rate and water cut, the 5 zone configuration has resulted in the best results. This may be due to the fact that the initial EnKF ensembles are not the same in these experiments.



(a) Water cut in PROD2 – Final Ensemble
5 zone configuration



(b) Water cut in PROD2 – Final Ensemble
100 zone configuration

Figure 7-17 Water cut for final ensembles (a) 5 zones configuration , (b) 100 zones configuration, in well PROD2. Results for each ensemble member are shown by the gray curves. Reference data are plotted in red. The black dashed line separates the history-matching period (0-516 days) from the forecasting period (517-1612 days).

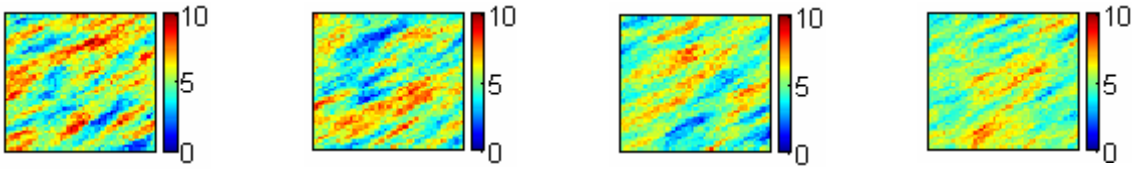
The $RMS_{prod,1}$ (equation 3.2) for well production data are given in Table 7-3. In order to make the comparison easier, the values for the final ensembles are normalized to the one obtained with only one zone. As can be observed, by increasing the number of zones in the reservoir, the RMS values decrease significantly, the best results being obtained with 100 zones.

Table 7-3 Effect of the local deformation with different zone numbers: Value of the normalized final $RMS_{prod,1}$ obtained by increasing the number of zones. Normalization is performed with respect to the final values for the global case (one zone).

Number of Zones	BHP	SOR	WCT
1	1.0	1.0	1.0
5	0.8049	0.6781	0.6770
100	0.1877	0.5770	0.5762

From this experiment, it seems that the larger the number of zones, the better the production data match. Indeed, the search space is larger. Thus, we are more likely to find reservoir models, which can regenerate the reference well production data.

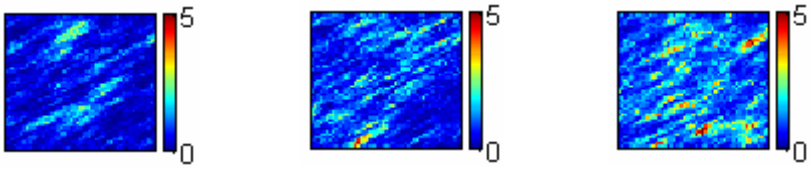
Figure 7-18 shows the mean model for the final EnKF ensembles for different zonations. Considering the global case (one zone) as the base case, the static RMS value $RMS_{stat,1}$ (equation 3.4) was reduced to 93% and 86% of the base value by increasing the number of zones to 5 and 100, respectively.



(a) Reference Model (b) 1 Zone (c) 5 Zones (d) 100 Zones

Figure 7-18 Final Mean model for the ln kh for local gradual deformation with several zones: (a) Reference model, (b) 1 zone, (c) 5 zones, (d) 100 zones.

In addition, Figure 7-19 shows that by increasing the number of zones, the variance decreases less during assimilations, making it more suitable for further prediction studies.



(a) 1 Zone (b) 5 Zones (c)100 Zones

Figure 7-19 Final Variance model for ln kh for local gradual deformation: (a) 1 zone, (b) 5 zones, (c) 100 zones

7.3.2 History-Matching with gradual deformation applied to the realizations for each property of interest

In this section, the assimilation described in section 7.3.1.2 is repeated, except that the gradual deformation (which is global here) is applied to the realizations of the log permeability rather than to the underlying Gaussian white noises. In the first experiment, the EnKF ensemble size is set to 50 and the background ensemble size to 100. Then, we further extend the size of the background ensemble to 1000 and 2500. Although not presented here, in all cases, the performance of the final ensemble was superior to that of the initial ensemble in terms of tracking the trend of the reference production data with less uncertainty. However, we encountered the same problems as previously, such as the nonlinearity due to the gradual deformation method. Thus, the traditional EnKF still performs better in terms of regenerating the reference well production data.

Table 7-4 Effect of the background ensemble size: value of the normalized final $RMS_{prod,2}$ obtained by increasing the background ensemble size. Normalization is performed with respect to the final values for the background ensemble of size 100.

Background Ensemble Size	BHP	SOR	WCT
100	1.0	1.0	1.0
1000	0.7452	0.8740	0.8757
2500	0.6211	0.7740	0.7740

Table 7-4 shows the final $RMS_{prod,1}$ (equation 3.2) for these three experiments, the results being normalized to the final RMS value obtained for the background ensemble of size 100. As shown previously, increasing the background ensemble size contributes to enhance the performance of the algorithm.

7.4 Conclusions

The objective of this chapter was to propose an algorithm to combine the EnKF with the gradual deformation parameterization method. The idea is to be able to preserve the properties of the prior (or initial) model throughout the assimilations with the EnKF. This is mainly important whenever geological constraints should be preserved throughout the optimization problem, for instance when working with facies models. The algorithm is based on the update of gradual deformation parameters rather than petrophysical properties as static variables in the EnKF state vector. These parameters are then applied to a background ensemble in order to generate petrophysical property realizations used for the forecast step. The algorithm was applied to the 2D synthetic case study introduced in Chapter 3. This may be explained by the limited search space for a set of constrained realizations and by the fact that gradual deformation increases the nonlinearity of the problem. In addition, the EnKF ensemble members are limited to those realizations which can be constructed by the gradual deformation parameterization. Hence, the method is more restrictive. The performance of the algorithm can be improved by either increasing the size of the background ensemble or by making use of local gradual deformations.

The research in this area can be further continued by considering iterative approaches within the existing algorithm or proposing new algorithms to combine the two methods. Also, application of the method to a more complicated case study, such as the PUNQ-S3 case introduced in Chapter 4, would be interesting. This would lead us to investigate the capability of the new algorithm in preventing some drawbacks of the EnKF such as parameter overestimation. The method could also be applied to a facies case study to investigate its capabilities in dealing with the non-Gaussian nature of facies models.

8 Concluding remarks and perspectives

The study of a petroleum reservoir is a multidisciplinary task integrating several sources of data to build the most representative reservoir model. Its main stages were provided in Chapter 1. The last step consists of constraining the model to dynamic data collected on the field during production: it is called “history-matching”. The research work in this thesis focused on a particular history-matching technique, the "Ensemble Kalman Filter (EnKF)". This sequential data assimilation method has recently become very popular in reservoir engineering. It makes it possible to integrate data as soon as they are acquired and provides a set of constrained models. The theoretical framework of ensemble Kalman filtering was detailed in Chapter 2, followed by a state of the art for petroleum applications. Most EnKF applications in this area are deemed to be successful. However, some fundamental assumptions of the method are not usually satisfied, such as the Gaussianity of state variables, the linear dependence between measurements and state variables, and the infinite size of the ensembles. This can lead to a deterioration of the filter’s performance, and several solutions were proposed in the literature as detailed in Chapter 2.

In this context, the objective of the thesis was two fold. First, we assess the performance of the EnKF for history-matching, considering several case studies and controlling parameters settings (ensemble size, assimilation time step, ...). We then attempted to investigate the merits and drawbacks of this conditioning technique. Second, new algorithms were developed that combine the EnKF with parameterization methods in order to preserve the spatial variability of the initial (or prior) models during assimilations.

More precisely, for handling the first objective, we applied the traditional EnKF to three synthetic case studies with increasing complexity. First, the log of permeability field of a 2D reservoir was constrained to production data (Chapter 3). Second, a variant of the PUNQ-S3 case study was defined, considering both porosity and permeability distributions as uncertain parameters (Chapter 4). Finally, the problem of non-continuous facies models was addressed in Chapter 5.

In each case, the spread between the production responses simulated with the ensemble members was reduced through assimilation, and most ensemble members were capable of

capturing the trend of the reference production data in both history-matching and prediction intervals. Also, the petrophysical property distributions in the final ensembles shared common features with the reference models. Using different controlling parameter settings for the EnKF on the first two cases, it was observed that the choice of the initial ensemble can strongly influence the results, especially for small ensembles. For the tests performed here, a size of 100 appeared appropriate for the ensembles, and only small improvements were obtained with larger ensembles. For smaller ensemble size, *i.e.* 50, the final variance within the updated ensemble decreased significantly, making it inappropriate for estimating covaraiances. In addition, small assimilation intervals may be required to capture abrupt changes in the flow behavior, induced for instance by changes in the flow rates or by wells closing and opening. The choice of the measurement uncertainty level is of less importance provided that measurements are treated as an ensemble. Finally, parameters overestimation could be observed, as a consequence of spurious correlations in the state matrix due to the small ensemble size. For facies models, facies proportions were not preserved.

The second part of this thesis introduced new algorithms, which aimed to perform assimilations with the EnKF while preserving the prior two-order statistics of the petrophysical properties distributions. They combine the EnKF with two parameterization techniques, the pilot point method (Chapter 6) and the gradual deformation method (Chapter 7). The basic idea is to generate a background ensemble of realizations that is perturbed after each assimilation from the updated values of related parameters: values at pilot points or gradual deformation parameters. The resulting algorithms were applied to the 2D case study introduced in Chapter 3.

When considering pilot points, the EnKF is used to update the values of petrophysical properties at the pilot point locations. The perturbed values are propagated to the whole field through kriging. The performance of this algorithm depends on the number and positions of the pilot points. As more pilot points are considered, well production data are better matched and more features of the reference permeability model are captured. However, the distance between pilot points should be smaller than one correlation length in order to preserve spatial variability. This requirement was violated to some degree in the best assimilation scenario in Chapter 6.

The algorithm combining the EnKF and the gradual deformation is more restrictive than the EnKF method. First, the search space is limited to the possible combinations of the background ensemble members. Second, the relation between gradual deformation parameters

and EnKF ensemble members is nonlinear. It is not sure if a set of updated gradual deformation parameters can result in a set of optimal petrophysical properties realizations. However, the performance of this algorithm can be improved by introducing more flexibility, either increasing the size of the background ensemble or making use of local deformations. Again, this method strongly depends on the choice of the initial gradual parameter values and background ensembles.

Several perspectives can be envisaged for future work, such as

- the application of the traditional EnKF method to a real field case study;
- the application of the traditional EnKF method with localization to the PUNQ-S3 cases study;
- the application of the traditional EnKF to non stationary facies proportions instead of the underlying standard Gaussian fields;
- the application of the algorithms combining the EnKF with the pilot point and gradual deformation methods to the PUNQ-S3 variant and facies case;
- the investigation of other possible combinations of EnKF and gradual deformation, introducing iterations for instance.

References

1. Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, and B., 2009, The Ensemble Kalman Filter in Reservoir Engineering – a Review, SPE Journal, v. 14, n. 3, p. 393-412.
2. Anderson, J.L., 2007, Exploring the Need for Localization in Ensemble Data Assimilation Using a Hierarchical Ensemble Filter, Physica D: Nonlinear Phenomena, v. 230, n. 1-2, p. 99-111.
3. Arroyo-Negrete, E.R., 2006, Continuous Reservoir Model Updating Using Streamline-Assisted Ensemble Kalman Filter, Paper SPE 106518 presented at the 2006 Annual Technical conference Exhibition, Texas, 24-27 September.
4. Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T., 2002, A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on signal processing, v. 50, n. 2.
5. Aziz, K., and Settari, A., 1979, Petroleum Reservoir Simulation: Applied Science, London.
6. Bertino, L., Evensen, G., and Wackernagel, H., 2003, Sequential Data Assimilation Techniques in Oceanography, International Statistical Review, v. 71, n. 2, p. 223-241.
7. Bianco, A., Cominelli, A., Dovera, L., Nævdal, G., and Vallès, B., 2007, History Matching and Production Forecast Uncertainty by Means of the Ensemble Kalman Filter: a Real Field Application. Paper SPE 107161, presented at the 2007 SPE/EAGE Annual Conference and Exhibition, United Kingdom, 11-14 June.
8. Bishop, C.H., and Hodyss, D., 2007, Flow-Adaptive Moderation of Spurious Ensemble Correlations and its Use in Ensemble-Based Data Assimilation, Quarterly Journal of Royal Meteorological Society, v. 133, p. 2029-2044.
9. Bissel, R.C., Dubrule, O., Lamy, P. Swaby, P., and Lepine, O., 1997, Combining geostatistical modeling with gradient information for history matching: The pilot point method, Society of Petroleum Engineers, 38730.
10. Burgers, G., van Leeuwen, and Evensen, G., 1998, On the analysis scheme in the ensemble Kalman filter, Monthly Weather Review, v. 126, p. 1719-1724.

11. Caers, J., 2003, History Matching Under Training-Image-Based Geological Model Constraints, SPE Journal, v. 8, p. 218-226.
12. Chavent, G., Dupuy, M., and Lemonnier, P., 1973, History-Matching by use of optimal control theory, SPE ATCE, 4627, Las Vegas, Nev., USA.
13. Chen., Y., Oliver, D.S., and Zhang, D., 2008, Efficient Ensemble-Based Closed Loop Production Optimization. Paper SPE 112873 presented at the 2008 Improved Recovery Symposium, Tulsa, 19-23 April.
14. Chen, Y., D.S. Oliver, and D. Zhang, 2009, Data assimilation for nonlinear problems by ensemble Kalman filter with reparameterization: Journal of Petroleum Science and Engineering, v. 66, p. 1-14.
15. Chilès, J.P., and Delfiner, P., 1999, Geostatistics – Modeling Spatial Uncertainty, Wiley series in probability and statistics, New York, USA.
16. Choe, J., and Park, K., 2006, Use of Ensemble Kalman Filter for 3-Dimensional Reservoir Characterization during Water flooding, Paper SPE 100178, presented at the 2006 68th EAGE Conference and Exhibition, Austria, 12-15 June.
17. Cohn, S., 1997, An introduction to estimation theory, Journal of the Meteorological Society of Japan, v. 75, p. 257–288.
18. Cossé, R., 1993, Basics of Reservoir Engineering, Gulf Publishing Company, Book Division, Houston, Texas, USA.
19. Cuypers, M., Dubrule, O., Lamy, P., and Bissel, R., 1998, Optima choice of inversion parameters for history matching with pilot point method, ECMOR VI, Peebles, UK.
20. de Marsily, G., Lavedan, G., Boucher, M., and Fasanino, G. 1994, Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. Geostatistics for Natural Resources Characterization, Part 2. Dordrecht: Reidel, p. 831-849.
21. Devegowda, D., Arroyo-Negrete, E., Datta-Gupta, A., Douma, S.G., 2007, Efficient and Robust Reservoir Model Updating Using Ensemble Kalman Filter with Sensitivity-Based Covariance Localization, Paper SPE 106144, presented at the 2007 Reservoir Simulation Symposium, Houston, 26-28 February.
22. Doucet, A., Freitas, N., and Gordon, N. (Ens.), 2001, Sequential Monte Carlo methods in Practice, Statistics for Engineering and Information Science, Springer-Verlag New York.
23. Dovera, L., and Della Rossa, E, 2007, Ensemble Kalman Filter for Gaussian Mixture Models. In Petroleum Geostatistics, 2007, 10-14 September 2007, Cascais, Portugal, A16, Utrecht, The Netherlands: Extended Abstracts Book, EAGE Publications BV.

24. Evensen, G., 1994, Sequential Data Assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Resources, 99:10, 143-10, 162.
25. Evensen, G., and van Leeuwen, P.J., 1996, Horizontal and Vertical Structure of the Representer Functions for Sea Structure Measurements in a Coastal Circulation Model, v. 30, p. 2627-2635.
26. Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53:343-367.
27. Evensen, G., Hove, J., Meisingset, H.C., Reiso, E., Seim, K.S., Espelid, Ø, 2007, Using the EnKF for Assisted History Matching of a North Sea Reservoir Model, Paper SPE 106184, presented at the 2007 Reservoir Simulation Symposium, Texas, 26-28 February.
28. Evensen, G., 2007, Data Assimilation: The Ensemble Kalman Filter. Berlin: Springer Verlag.
29. Floris, F.J.T., Bush, M.D., Cuypers, M., Roggero, F., and Syversveen, A.R., 2001, Methods for quantifying the uncertainty of production forecasts: a comparison study, Petroleum Geoscience, v. 7, p. 87-96.
30. Furrer, R., and Bengtsson, T., 2007, Estimation of High Dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants. Journal of Multivariate Analysis, v. 98, n. 2, p. 227-255.
31. Galli, A., Beucher, H., Le Loc'h, G., Doligez, B., and Heresim group., 1993, The Pros and Cons of the Truncated Gaussian Method, Geostatistical Simulations, ed. Armstrong and P.A. Dowd, 217-233. Dordrecht, The Netherlands: Quantitative Geology and Geostatistics, Kluwer Academic Publishers.
32. Gao, G., Zafari, M., and Reynolds, A.C., 2005, Quantifying the uncertainty for the PUNQ-S3 problem in a Bayesian setting with the RML and EnKF, SPE reservoir simulation symposium (SPE 93324).
33. Gaspari, G., and Cohn, S.E., 1999, Construction of Correlation Functions in Two and Three Dimensions, Quarterly Journal of Royal Meteorological Society, v. 125, p. 723-757.
34. Gelb, A., 1974, Applied Optimal Estimation, MIT Press Cambridge.
35. Gómez-Hernández, J.J., Sahuquillo, A., and Capilla, José, E., 1997, Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Pizometric Data, 1. Theory, Journal of Hydrology, v. 203, n. 3.

36. Gu, Y., and Oliver, D.S., 2005, History Matching of the PUNQ-S3 Reservoir Model Using the Ensemble Kalman Filter: Journal of Petroleum Science and Engineering, v. 147, p. 147-161.
37. Gu, Y. and Oliver, D.S., 2006, The Ensemble Kalman Filter for Continuous Updating of Reservoir Simulation Models, Journal of Energy Resources Technology, v. 128, n. 1, p. 79-87.
38. Gu, Y., and Oliver, D.S., 2007, An Iterative Ensemble Kalman Filter for multiphase fluid flow data assimilation, SPE Journal, v., 12, n. 4, p. 438-446.
39. Hacker, J.P., Anderson, J.L., and Pagowski, M., 2007, Improved Vertical Covariance Estimate for Ensemble-Filter Assimilation of Near-Surface Observations. Monthly Weather Review, v. 135, n. 3, p. 1021-1036.
40. Hamill, T.M., Whitaker, J.S., and Snyder, C., 2001, Distance-Dependent Filtering of Background Error Covariance Estimate in an Ensemble Kalman Filter, Monthly Weather Review, v. 129, n. 11, p. 2776-2790.
41. Haugen, V.E.J., and Evensen, G., 2002, Assimilation of SLA and SST data into OGCM for the Indian Ocean: Ocean Dynamics, v. 52, p. 133-151.
42. Haugen, V.E.J., Natvik, L., Evensen, G., Berg, A., Flornes, K., and Nævdal, G., 2006, History Matching Using the Ensemble Kalman Filter on a North Sea Field Case, Paper SPE 102430, presented at the Annual Technical Conference and Exhibition, Texas, 24-27 September.
43. Heidari, L., Gervais, V., Le Ravalec, M., and Wackernagel, H., History Matching of Reservoir Models by the Ensemble Kalman Filtering (EnKF): Review of the State of the Art and a Sensitivity Study, American Association of Petroleum Geologists (AAPG), 2009.
44. Houtekamer, P.L., and Mitchell, H.L., 1998, Data Assimilation Using an Ensemble Kalman Technique, Monthly Weather Review, v. 126, n. 6, p. 796-811.
45. Houtekamer, P.L., and Mitchell, H.L., 2001, A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. Monthly Weather Review, v. 129, n. 1, p. 123-137.
46. Hu, L.-Y., 2000, Gradual deformation and iterative calibration of Gaussian-related stochastic models, Mathematical Geology, vol. 32, n. 1, p. 87-108.
47. Hu, L.-Y., Blanc, G., Noetinger, B., 2001, Gradual Deformation and Iterative Calibration of Sequantial Stochastic Simulations, Mathematical Geology, vol. 33, n. 4, p. 475-489.
48. Hu, L.-Y., 2003, History Matching of Object-Based Stochastic Reservoir Models, Middle East Oil Show, 9-12 June, Bahrain.

49. Indelman, P., and Dagan, G., 1993, Upscaling of permeability of anisotropic heterogeneous formations, 1- The general framework, Water Resources Research, v. 29, n. 4, p. 917-923.
50. Jafarpour, B., and McLaughlin, D.B., 2007, History Matching with an ensemble Kalman Filter and Discrete Cosine Parameterization. Paper SPE 108761, presented at the 2007 Annual Technical Conference and Exhibition, California, 11-14 November.
51. Jazwinski, A. H., 1970, Stochastic Processes and Filtering Theory, Academic Press, San Diego, California.
52. Kalman, R., 1960, A new approach to linear filtering and prediction problems, Transactions of ASME, Journal of Basic Engineering, v. 82, p. 35-45.
53. Le Gallo, Y., and Le Ravalec-Dupin, M., 2000, History Matching Geostatistical Reservoir Models with Gradual Deformation Method, SPE Annual Technical Conference and Exhibition, 1-4 October, Dallas, Texas.
54. Le Loc'h, G., and Galli., 1997, Truncated pluriGaussian method: Theoretical and practical points of view, in Geostatistics Wollogong'96, v. 1, p. 211-222.
55. Le Ravalec, M., Noetinger, B., and Hu, L.Y., 2000, The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulation, Mathematical Geology, v. 32, n. 6, p. 701-723.
56. Le Ravalec-Dupin, M., Noetinger, B., Hu, L.-Y., and Blanc, G., 2001, Conditioning to dynamic data: an improved zonation approach, Petroleum Geosciences, v. 7, p. S9-S16.
57. Le Ravalec-Dupin, M., 2005, Inverse Stochastic Modeling of Flow in Porous Media. Paris: IFP Publications.
58. Le Ravalec-Dupin, M. and Hu, L.Y., 2007, Combining the Pilot Point and the Gradual Deformation Methods for Calibrating Permeability Models to Dynamic Data, Oil and Gas Science and Technology, Rev. IFP Energies nouvells, v. 62, n. 2, p. 169-180.
59. La Venue, A.M. , and Pickens, J.F., 1992, Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resources Research, v. 28, n. 6, p. 1543-1569.
60. La Venue, A.M., RamaRao, B.S., de Marsily, G., and Marietta, G., 1995, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 2. Application, Water Resources Research, v.31, n.3, p.495-516.
61. Li, G. and Reynolds, A.C., 2007, An Iterative Ensemble Kalman Filter for Data Assimilation. Paper SPE 109808 presented at the 2007 Annual Technical Conference Exhibition, California, 11-14 November.

62. Liang, B., Alpak, F.O., Sepehrnoori, K., and Delshad, M., 2007, A Singular Evolutive Interpolated Kalman Filter for Rapid Uncertainty Quantification, Paper SPE 106170, presented at the 2007 Reservoir Simulation Symposium, Houston, 26-28 February.
63. Liu, N., and Oliver, D.S., 2005, Critical Evaluation of the Ensemble Kalman Filter on History Matching of Geologic Facies, Journal of Petroleum Science and Engineering, v. 47, n. 6, p. 470-477.
64. Lorenc, A., 1986, Analysis methods for numerical weather prediction. Quarterly Journal of Royal Meteorological Society, v. 112, p. 1177-1194.
65. Lorenc, A.C., 2003, The Potential of the Ensemble Kalman Filter for NWP – A Comparison with the 4D-Var, Quarterly Journal of Royal Meteorological Society, v. 129, p. 3183-3203.
66. Lorentzen, R.J., Fjelde, K.K., Frøyen, J., Lage, A.C.V.M., Nævdal, G., and Vefring, E.H., 2001, Underbalanced and Low-Head Drilling Operations: Real Time Interpretation of Measured Data and Operational Support. Paper SPE 71384 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 30 September-3 October.
67. Lorentzen, R.J., Nævdal, G., and Lage, A.C.V.M., 2003, Tuning of Parameters in a Two-Phase Flow Model Using an Ensemble Kalman Filter. International Journal of Multiphase Flow, v. 29, n. 8, p. 1283-1309.
68. Lorentzen, R.J., Nævdal, G., Valles, B., Berg, A.N., and Grimstad, A.A., 2005, Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models, SPE 96375.
69. Margulis, S., A., McLaughlin, D., Entekhabi, D., and Dunne, S., 2002, Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, Water Resources, v. 38, n. 12, p. 35.1-35.18.
70. McLaughlin, D., and Townley, L.R., 1996, A reassessment of the groundwater inverse problem, Water Resources Research, v.32, n.5 , p. 1131-1161.
71. Mezghani, M., Roggero, F., and Raymond, J.P., 1998, Emplacement optimal des points pilotes pour le calage des données dynamique de tests de puits, Rapport IFP 45154.
72. Moreno, D., and Aanonsen, S.I., 2007, Stochastic Facies Modeling Using the Level Set Method, In Petroleum Geostatistics, 2007, 10-14 September, Cascais, Portugal, A18. Utrecht, The Netherlands: Extended Abstract Book, EAGE Publications BV.
73. Nævdal, G., Mannseth, T., and Vefring, E.H., 2002, Near-well reservoir monitoring through Ensemble Kalman filter, SPE/DOE Improved Oil Recovery Symposium, Tulsa, Oklahoma, SPE 75235

74. Nævdal, G., Johnsen, L.M., Aanonsen, S.I., and Vefring, E.H., 2003, Reservoir monitoring and continuous model updating using ensemble Kalman filter, SPE 84372.
75. Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalany, E., Patil, D.J., and Yorke, J.A., 2004, A Local Ensemble Kalman Filter For Atmospheric Data Assimilation, Tellus A, v. 56, n. 5, p. 415-428.
76. Oliver, D.S., He, N. and Reynolds, A.C., 1996, Conditioning Permeability Fileds to Pressure Data , European Conference for Mathematics of Oil Recovery
77. Press, W.H., Teufolsky, S.A., Vetterling, W.T., and Flannery, B.P., 1998, Numerical recipes in C, Cambridge University Press, New York, NY, USA.
78. PumaFlow, 2007, Reference Manual Release 2.0. September. BeicipFranlab –IFP.
79. RamaRao, B.S., La Venue, A.M., de Marsily, G., and Marietta, G., 1995, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1. Theory and computational experiments, Water Resources Research, v.31, n.3, p. 475-493.
80. Reynolds, A.C., Zafari, M., and Li, G., 2006, Iterative Forms of the Ensemble Kalman Filter, Proceeding of the 10th European Conference on the Mathematics of the Oil Recovery, Amsterdam, 4-7 September, Paper A030.
81. Roggero, F., Ding, D.Y., Berthet, Ph., Lerat, O., Cap, J. and Schreiber, E., 2007, Matching of Production History and 4D Seismic Data – Application to the Girassol Field, Offshore Angola, Paper SPE 109929, Presented at the SPE ATC&E, Anaheim, CA, 11-14 Nov.
82. Rubin, Y., and Gomez-Hernandez, J., 1990, A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations, Water Resources Research, v. 26, n. 4 , 691-701.
83. Seiler, A., Evensen, G., Skjervheim, J.-A., Hove J., and Vabø, J.G., 2009, Advanced Reservoir Management Workflow Using an EnKF Based Assisted History Matching Method, SPE 118906. This paper was prepared for presentation at the 2009 SPE Reservoir Simulation Symposium held in The Woodlands, Texas, USA, 2-4 February 2009.
84. Skjervheim, J.-A., Evensen, G., Aanonsen, S.I., Rund, B.O., Johansen, T.A., 2005, Incorporating 4D Seismic Data in Reservoir Simulation Methods Using Ensemble Kalman Filter, SPE 95789.
85. Sun, N.-Z., 1994, Inverse problems in groundwater modeling, Kluwer Academic Publishers, The Netherlands.
86. Tarantola, A., 1987, Inverse problem theory – Methods for data fitting and model parameter estimation, Elsevier Science Publishers, Amsterdam, Netherlands.

87. Thulin, K., G. Naevdal, and S.I. Aanonsen, 2008, Quantifying Monte Carlo Uncertainty in the Ensemble Kalman Filter: In proceedings of the 11th European Conference on the Mathematics of Oil recovery, Bergen, Norway, 8-11 September.
88. Vabø, J.G., G. Evensen, J. Hove, and J.A. Skjervheim, 2008, Using the EnKF with Kernel Methods for Estimation of Non-Gaussian Variables: 11th European Conference on the Mathematics of Oil Recovery.
89. Wang, C., Li., G., and Reynolds, A.C., 2007, Production optimization in Closed Loop Reservoir Management. Paper SPE 109805, presented at the 2007 Annual Technical Conference and Exhibition, California, 11-14 November.
90. Wen, X.-H., and Chen, W.H., 2005, Some Practical Issues on Real Time Reservoir Model Updating Using Ensemble Kalman Filter. Paper IPTC 10024, presented at the 2005 International Petroleum Technology Conference, Qatar, 21-23 November.
91. Wunsch, C., 1996, The Ocean Circulation Inverse Problem, Cambridge University Press, 442 pp.
92. Xue, G., and Datta-Gupta, A., 1997, Structure preserving inversion: an efficient approach to conditioning stochastic reservoir models to dynamic data, Society of Petroleum Engineers, 38727.
93. Zafari, M., and Reynolds, A.C., 2005, Assessing the Uncertainty in Reservoir Description and Performance Prediction with the Ensemble Kalman Filter, Paper SPE 95750, presented at the 2005 Annual Technical Conference and Exhibition, Texas, 9-12 October.
94. Zabalza-Mezghani, I., Mezghani, M. and Blanc, G., 2001, Constraining Reservoir Facies Models to Dynamic Data – Impact of Spatial Distribution Uncertainty on Production Forecasts, The 2001 SPE Annual Technical Conference and Exhibition, 30 September-3 October, New Orleans, Louisiana.

Appendix A Real Gas Storage Case Study

An effective solution for meeting the energy demand at high consumption seasons is gas storage. Gas can be bought at a lower price during the low consumption seasons, stored and then extracted during the high consumption season. Three main natural storage sites are: depleted gas reservoirs, porous aquifer formations and salt cavern reservoirs. In the first two cases, an accurate understanding of the petrophysical properties, *e.g.* porosity and permeability, of the underlying formation is required.

In this study, we aim to apply the EnKF for history-matching of well pressure data for a real underground gas storage (UGS) site based on a porous aquifer formation described by Le Ravalec-Dupin *et al.* (2004). The results will provide us with a more accurate understanding of the petrophysical properties distribution, *i.e.* porosity and permeability.

A.1 Case study description

A.1.1 Aquifer formation

An aquifer formation is a porous and permeable sand, sandstone or carbonate rock situated at depths of 1500 to 6000 ft and initially filled with water. This formation is a suitable site for gas storage, even in large quantities. As this site does not contain gas initially, a cushion gas has to be injected to maintain adequate pressure in the site for meeting gas rates during the extraction periods. This gas is not recoverable and its volume is comparable with the total storage capacity of the site. Any storage planning depends on the petrophysical characteristics of the porous rock such as porosity and permeability. Before the beginning of storage, these properties are known with a high level of uncertainty as the site has never been used for petroleum production purposes. As soon as the storage plan starts, the obtained dynamic data through wells can be used in a history-matching scenario to improve the understanding of the underground storage site.
Our case study is a UGS site located in an aquifer formation close to Paris, France and operated by GDF Suez. The main uncertain petrophysical properties are porosity and permeability.

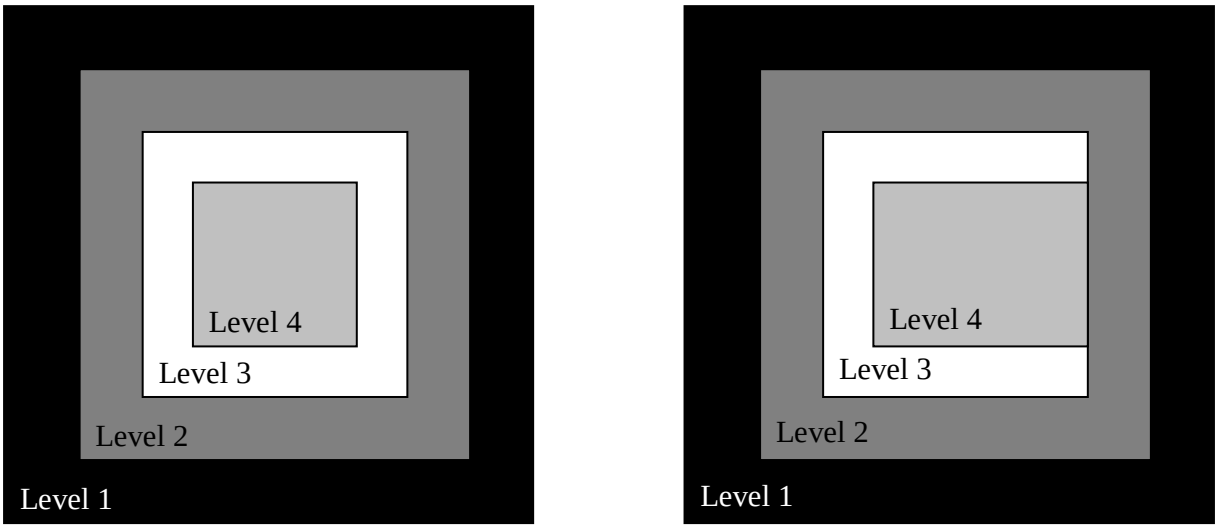
A.1.2 Gas storage planning

The storage plan consists of two main periods: a filling period with cushion gas injection and a cycling period consisting of gas injection during summer and gas withdrawal it during winter. In this case study, the first period lasts 7 years and the second one 10 years. The data collected during the first period, *i.e.* cushion gas injection, are used to perform history-matching. The data collected during the second period can be used to assess the quality of predictions.

A.1.3 Numerical reservoir model

The numerical reservoir model is based on an embedded grid system. This grid setting described in Figure A-1 Description of grid system for the numerical reservoir model

(a), consists of 4 grid levels with decreasing grid block sizes from Level 1 to 4.



(a) Initial numerical reservoir model **(b) Final numerical reservoir model**

Figure A-1 Description of grid system for the numerical reservoir model

Level 1 contains the coarsest grid blocks: there are a total of 25×23×2 grid blocks with a size of 13.5 km in the horizontal direction and a varying thickness through each layer. The finest grid blocks are those of Level 4 with a size of 750 m in the horizontal direction. Initially, this level consists of 33×21×9 grid blocks with varying thickness throughout layers. However, some of the observation wells are placed in Level 3. The petrophysical properties of grids around them can be improved by the history-matching procedure. Therefore, we have downscaled them and included them in Level 4, as can be seen in Figure A-1 Description of grid system for the numerical reservoir model

(b)). Therefore, in the rest of this study, Level 4 consists of 39×21×9 grid blocks with a size of 750 m in the horizontal direction.

Gas storage is performed in the 7 intermediate layers, *i.e.* L2-L8, of Level 4. Thus, in this study, we try to improve the petrophysical property descriptions in these layers only. The properties of layers L1 and L9 in Level 4 and those of Levels 1 to 3 remain unchanged during the history matching procedure. As a whole, a total of 4888 active cells are considered.

A.1.4 Petrophysical properties

Petrophysical properties of Levels 1 to 3 and those in layers 1 and 9 of Level 4 in are assumed to be known without uncertainty based on some prior knowledge on this site. For those of layers 2 to 8 in Level 4, static well data are assumed to be known with a certain level of uncertainty. Processing those data provides the following relationship between porosity and permeability in the different layers:

$$\log_{10}(k) = A\varphi + B \qquad \text{A.1.}$$

A and B are constants, k and φ are horizontal permeability and porosity, respectively. The mean and variance for the porosity fields as well as constants A and B for porosity-permeability correlations are reported in Table A-1 .

Table A-1 Petrophysical properties of layers 2-8 in Level 4. Porosity is defined in [0,1]. Permeability is given in mD.

Layer	Porosity mean	Porosity variance	A	B
L2	0.2292	0.00278	6.8729	+0.4043
L3	0.1763	0.00608	9.5409	-0.3517
L4	0.0333	0.00177	16.967	-0.8234
L5	0.1895	0.00441	11.257	-0.6291
L6	0.1637	0.00545	9.2728	+0.1777
L7	0.1906	0.00334	7.6666	+0.0679
L8	0.1834	0.00556	9.4154	+0.5935

The permeability hard data at well locations associate a spherical and isotropic variogram to the horizontal permeability fields; however the variogram model for porosity is unknown. Therefore, it was assumed that porosity fields have the same variogram model as that of the permeability fields. The correlation length was set to 250 m for all layers and the ratio of vertical to horizontal permeability to 0.1 in all layers. The realizations for porosity in the different layers are generated by the FFT-MA algorithm (Le Ravalec *et al.*, 2000).

A.1.5 Well placement

Gas injection and production scenarios are performed from several wells serving both as injectors and producers during cushion gas injection period and filling period. These wells are indicated by white circles in Figure A-2 Well placement in the reservoir (Level 4). Observation wells are displayed by cross and injection/injection wells by white circles (from Le Ravalec-Dupin *et al.* (2004))

. However, during these two periods, the reservoir pressure in Level 4 is recorded by 12 observation wells (W1-W12) placed all around Level 4 and indicated by crosses in Figure A-2 Well placement in the reservoir (Level 4). Observation wells are displayed by cross and injection/injection wells by white circles (from Le Ravalec-Dupin *et al.* (2004))

.

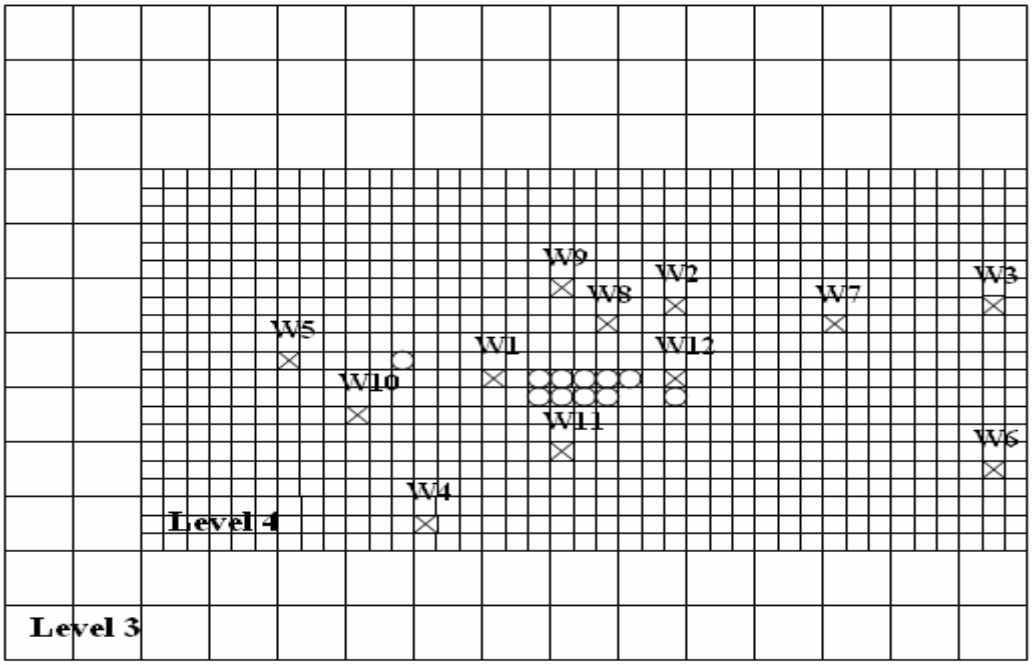


Figure A-2 Well placement in the reservoir (Level 4). Observation wells are displayed by cross and injection/injection wells by white circles (from Le Ravalec-Dupin *et al.* (2004))

Pressure data are collected at irregular intervals at observation wells and are used to calibrate porosity and permeability fields in the 7 intermediate layers of Level 4 during the history-matching procedure.

A.2 History matching on measured data with the EnKF

During the cushion gas injection period, injection and production rates (through the corresponding wells) undergo several changes. Data are gathered at irregular intervals with respect to these changes in the flow rates (Figure A.3).

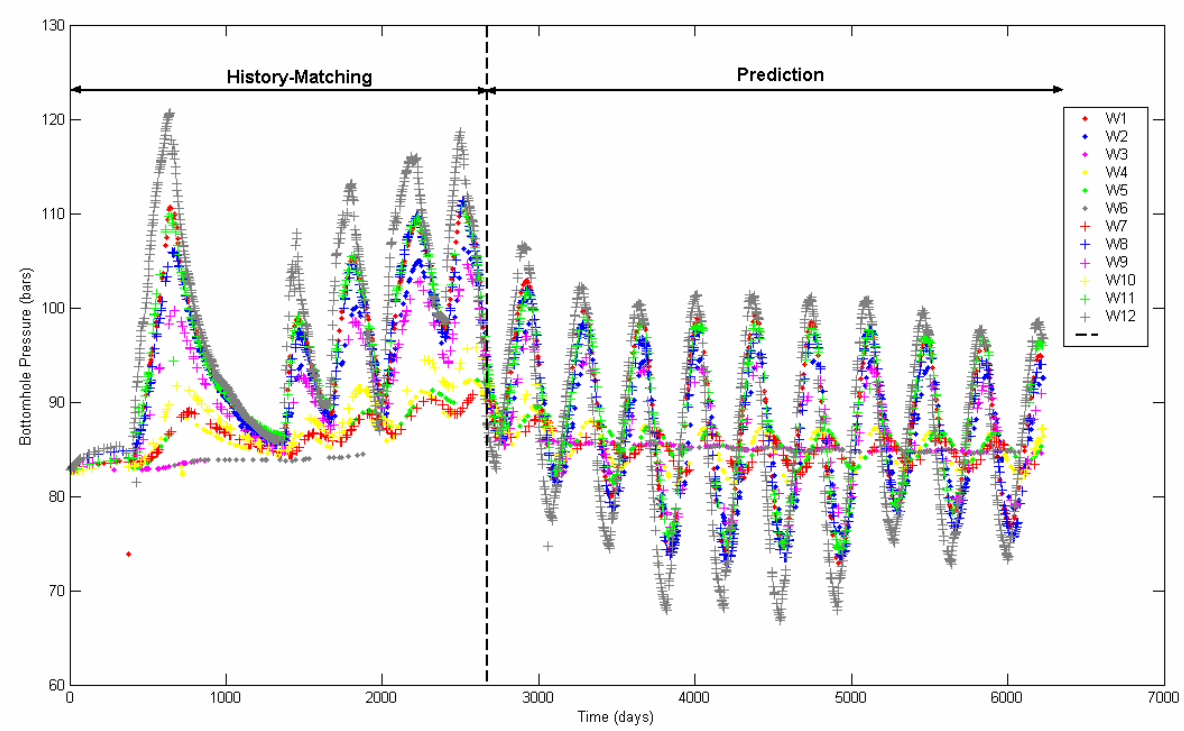


Figure A-3 Observed pressure during the history-matching (cushion gas injections) period and (cycling period).

A.2.1 Assimilation steps with the EnKF

The cushion gas injection period (or history-matching period) lasts 2671 days. Several scenarios were considered to obtain the best assimilation scenario and. Two of these are presented below:

- First scenario

The history-matching interval was divided to 89 successive intervals of one month. Then, at the end of each month, the EnKF is used to assimilate the average pressure observed at each observation well during that month.

- Second scenario

During the first 444 days of the history-matching period, gas is injected in the reservoir. From 445 days on, some wells start to produce gas while the others continue to inject it. In this scenario, average pressures are assimilated at the end of each month during 444 days, leading to a total of 15 assimilation steps. Then, assimilation is performed whenever a new data is obtained at one of the observation wells leading to a total of 714 assimilation steps. Therefore, a total of 729 assimilations steps are considered in this scenario.

A previous history-matching experiment on the same field (Le Ravalec-Dupin *et al.*, 2004) resulted in a set of improved values for each of the mentioned properties. The improved ones are reported in Table A. 2. In the experiments which follow, we aim to investigate if this improved data set can result in better assimilation results by the EnKF.

Table A-2 Improved petrophysical properties of layers 2-8 in Level 4. Porosity is defined in [0,1] and permeability is given in mD.

Layer	Porosity mean	Porosity variance	A	B
L2	0.2410	0.00278	6.9258	+0.4021
L3	0.1731	0.00608	9.5123	-0.3568
L4	0.0304	0.00177	16.9869	-0.8326
L5	0.1793	0.00441	11.2694	-0.6430
L6	0.1797	0.00545	9.3238	+0.1748
L7	0.1943	0.00334	7.6402	+0.0562
L8	0.2042	0.00556	10.7611	+0.6381

In the following section, we perform three experiments:

- Experiment 1: First scenario + Initial properties reported in Table A. 1.
- Experiment 2: Second scenario + Initial properties reported in Table A. 1.
- Experiment 3: Second scenario + Improved properties reported in Table A. 2.

A.2.2 EnKF setting for the history-matching experiments

The state vector for each ensemble member j is composed of:

- The transformed porosity values (based on transformations introduced in section 4.2) at each active grid cell (static variable): $trans(\varphi_{j,i})$, $i = 1,...,Active\,Cells$;
- The transformed permeability values (based on transformations introduced in section 4.2) at each active grid cell (static variable): $trans(k_{j,i})$, $i = 1,...,Active\,Cells$;
- Pressure at all active grid cells (dynamic variable): $P_{j,i}$, $i = 1,...,Active\,Cells$;
- Water saturation at all active grid cells (dynamic variable): $Sw_{j,i}$, $i = 1,...,Active\,Cells$;
- Bottom hole pressure at observation wells: $BHP_{j,k}$, $k = 1,...,12$;

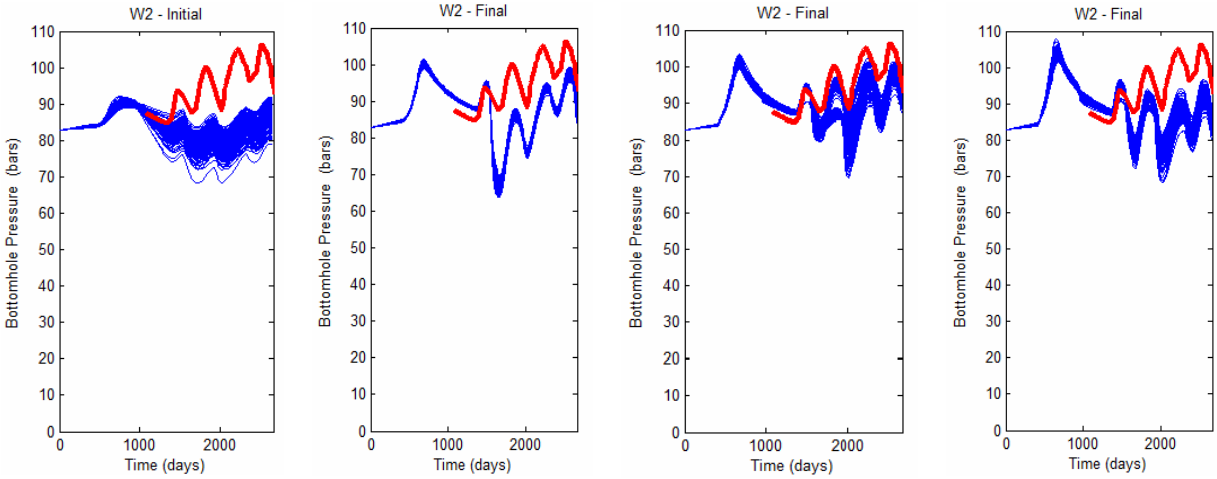
Therefore, the state vector for the j[th] ensemble member is defined by:

$$\psi_j = [trans(\varphi_{j,1}),...,trans(\varphi_{j,ActiveCells}),trans(k_{j,1}),...,trans(k_{j,ActiveCells}),P_{j,1},...,P_{j,ActiveCells}, \quad \text{A.2.}$$
$$Sw_{j,1},...,Sw_{j,ActiveCells},BHP_{j,1},...,BHP_{j,12}]$$

The uncertainty level for pressure measurements is kept to 5 bars and the ensemble size is set to 200.

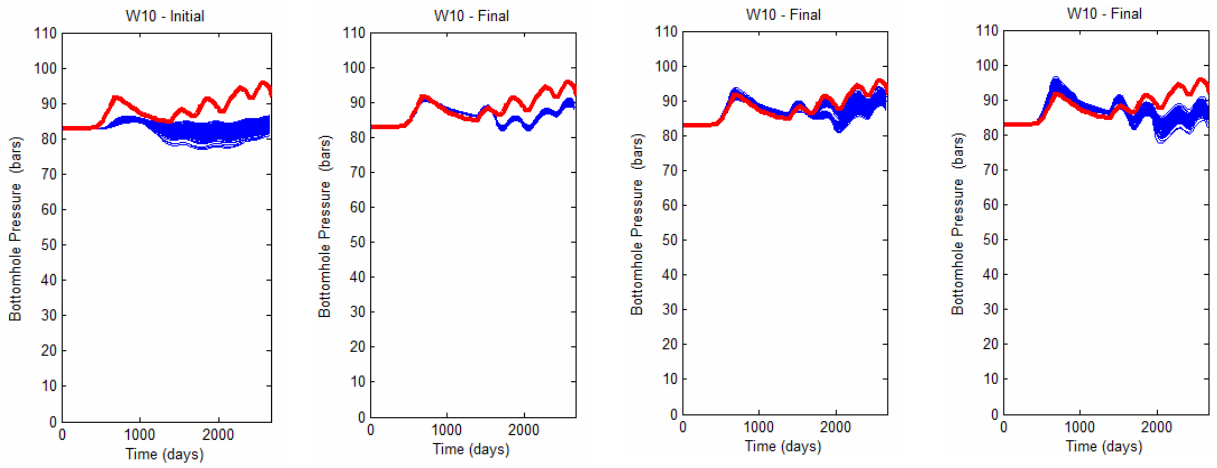
A.3 History-Matching results with the EnKF assimilations

The initial ensemble is the same for experiments 1 and 2. The initial ensemble for Experiment 3 is not represented in this section; however, its behavior is not that different from the one of experiments 1 and 2). In this section, the history-matching results are represented for three wells: W2, W10 and W12. The results for the other wells follow the same trend.



(a) Initial models Experiments 1 and 2 **(b) Experiment 1 – Final** **(c) Experiment 2 – Final** **(d) Experiment 3 – Final**

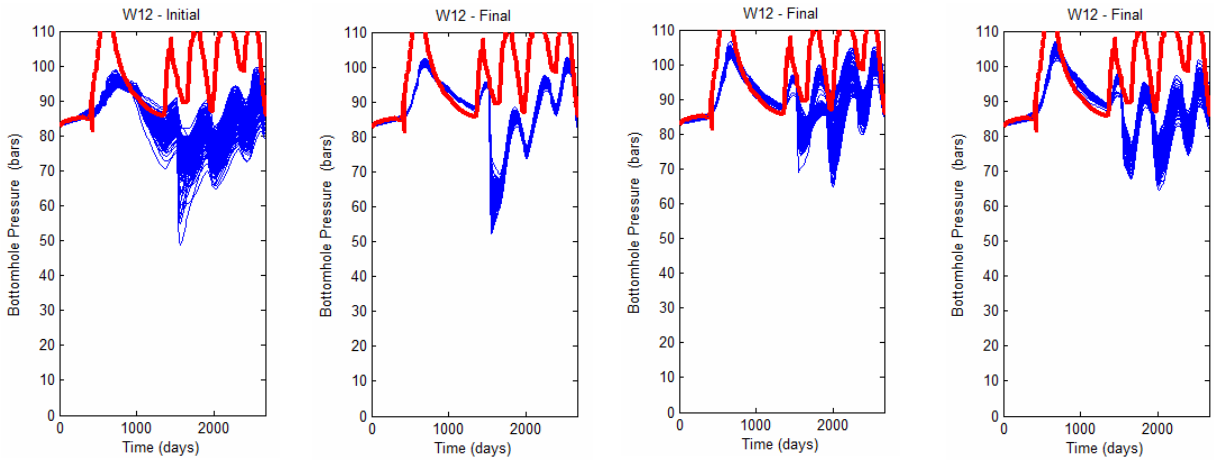
Figure A-4 Bottom hole pressure for well W2 during the history-matching period. (a) Initial ensemble for experiments 1 and 2, (b) Final ensemble for experiment 1, (c) Final ensemble for experiment 2, (d) Final ensemble for experiment 3. Pressure for each ensemble member is represented by the blue curves; reference data are plotted in red.



(a) Initial models Experiments 1 and 2 **(b) Experiment 1 – Final** **(c) Experiment 2 - Final** **(d) Experiment 3 – Final**

Figure A-5 Bottom hole pressure for well W10 during the history-matching period. (a) Initial ensemble for experiments 1 and 2, (b) Final ensemble for experiment 1, (c) Final ensemble for experiment 2, (d) Final ensemble for experiment 3. Pressure for each ensemble member is represented by the blue curves; reference data are plotted in red.

As can be observed in Figure A. 4 to A. 6, the initial models are all very different from the reference model. Although not presented here, we generated several initial ensembles but none of them was able to simulate values closer to the reference well pressure data. Le Ravalec-Dupin *et al.* (2004) generated a fine scale reservoir model and then upscaled it to the coarse scale model to perform the fluid flow simulation. Here, we generate the reservoir model over the coarse scale only. This may explain the behavior observed for the initial ensemble members.



(a) Initial models Experiments 1 and 2 **(b) Experiment 1 – Final** **(c) Experiment 2 – Final** **(d) Experiment 3 – Final**

Figure A-6 Bottom hole pressure for well W12 during the history-matching period. (a) Initial ensemble for experiments 1 and 2, (b) Final ensemble for Experiment 1, (c) Final ensemble for experiment 2, (d) Final ensemble for experiment 3. Pressure for each ensemble member is represented by the blue curves; reference data are plotted in red.

The well pressure data obtained from rerunning the fluid flow simulation with the final ensemble members for Experiment 1 are too far from the reference model. Moreover, the variation between ensemble members has decreased, indicating that assimilation resulted in several updated porosity and permeability fields by far different from the real values. This can be the consequence of the assimilation steps used in the study.

Well pressure data are better matched with the second assimilation scenario (Experiments 2 and 3). Although the match is not perfect, at least in Experiment 2 there are several ensemble members that can follow the trend of the reference pressure data within the predefined set of measurement errors. However, the match for well W12 is not that satisfactory. The reason can be the close placement of this well to the injectors/producers. As the injection and production schedule undergoes several changes, matching the data for wells close to these locations can be more difficult.

The first set of data for petrophysical properties (Table A. 1) have shown to provide better results compared to the second one (Table A. 2). This may seem contradictory at first place. However, as a history-matching problem is ill-posed, it is expected that several different models, significantly different, can reproduce the reference pressure data equally well. Moreover, it was observed already that history-matching with EnKF and some variational history-matching methods, such as the one used by Le Ravalec-Dupin *et al.* (2004), may not provide the same final matched models.

A.4 Conclusion

Matching the well production data for a real field is a challenging task. The production schedule may have several changes in injection/production rates and even opening/shutting of wells. Moreover, the initial petrophysical properties' distributions are built upon measurements collected at wells. If these data are not precise enough, the initial models are not credible enough.

In this case study, we applied the traditional EnKF method to a real Underground Gas Storage (UGS) site in order to match the well pressure data obtained at 12 observation wells. The uncertain petrophysical properties are porosity and permeability in 7 layers. Several scenarios were considered to perform history-matching with varying the assimilation steps and initial petrophysical statistical properties.

As the production schedule undergoes several changes, the second assimilation scenario was more successful in terms of matching pressure data. With this scenario, one assimilation was performed whenever a new data was collected at the field. Moreover, the best data set used for generating the initial porosity and permeability fields is the one derived directly from the hard data at well locations.

Other assimilations would be considered to improve the match for production data. This can be achieved by:

- selecting initial ensemble members closer to the reference pressure data either by better describing porosity fields (mean and variance) and porosity-permeability correlations, or by generating porosity and permeability fields on the fine scales before upscaling them to the coarser one;

considering the porosity-permeability coefficients as uncertain parameters in the state vector of the EnKF.